AFRL-HE-AZ-TR-2002-0004

UNITED STATES AIR FORCE RESEARCH LABORATORY

SKILL DECAY: A COMPARATIVE ASSESSMENT OF TRAINING PROTOCOLS AND INDIVIDUAL DIFFERENCES IN THE LOSS AND REACQUISITION OF COMPLEX SKILLS

Winfred Arthur Jr.

2828 Bishops Gate Circle
Bryan, TX 77807-4808

Winston Bennett Jr.

Air Force Research Laboratory
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

Eric A. Day

Theresa L. McNelly

December 2002

20030917 116

AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

NOTICES

Publication of this report does not constitute approval or disapproval of the ideas or findings. It is published in the interest of STINFO exchange.

The views expressed in this report are those of the authors and do not necessarily reflect official views of the US Air Force or the Department of Defense.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

WINSTON BENNETT JR.
Project Scientist

DEE H. ANDREWS
Technical Advisor

CURTIS J. PAPKE, Colonel, USAF
Chief, Warfighter Training Research Division

Copies of this paper may be requested from:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218
http://stinet.dtic.mil

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
December 2002	Final	Jan 1997 to July 1997

4. TITLE AND SUBTITLE

Skill Decay: A Comparative Assessment of Training Protocols and Individual Differences in the Loss and Reacquisition of Complex Skills

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62205F

6. AUTHOR(S)
Winfred Arthur Jr.
Winston Bennett Jr.
Eric A. Day
Theresa L. McNelly

5d. PROJECT NUMBER
1123

5e. TASK NUMBER
A2

5f. WORK UNIT NUMBER
19

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

2828 Bishops Gate Circle
Bryan, TX 77807-4808

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TR-2002-0004

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Air Force Research Laboratory Technical Monitor: Dr. Winston Bennett, (480) 988-6561 Ext. 297, DSN 474-6297

14. ABSTRACT

It has been consistently demonstrated that the active interlocked modeling (AIM) protocol, a dyadic training protocol developed by Shebilske, Regian, Arthur, and Jordan (1992), achieves a 100% increase in training efficiency over a control individual-trainee based condition. The purpose of the present study was to replicate this finding as well as to investigate its robustness in terms of skill loss and reacquisition. Thus, we compared the AIM-dyad training protocol to a standard individual protocol regarding the acquisition of a complex skill (i.e., Space Fortress; cf. Mane & Donchin, 1989). More importantly, skill loss after an 8-week non-practice interval and skill re-acquisition was also investigated. Despite half as much hands-on practice, the performance of dyadic trainees did not differ from that of individuals on tests of skill acquisition, loss, and re-acquisition. These findings provide strong support and justification for the ongoing use of innovative dyadic protocols for the training of pilots and navigators in both military an non-military settings.

15. SUBJECT TERMS
Criterion Development; Performance Appraisal; Ratings; Skill Decay; Training Assessment; Training Effectiveness; Training Efficiency; Training Evaluation; Training Needs Analysis; Training Transfer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Ms Liz Casey
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	104	480.988.6561 x-188 DSN 474-6188

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Skill loss refers to the loss of trained or acquired skills (or knowledge) after periods of non-use and is particularly salient and problematic in situations where individuals receive initial training on skills and knowledge which they may not be required to use or exercise for extended periods of time. Reserve personnel in the military, for example, may be provided formal training only once or twice a year with the expectation (or hope) that if these personnel are called up for active duty, they will need only a limited amount of refresher training to reacquire any skill that has been lost. Likewise disaster teams may go years without evacuating residents from affected areas, managing evacuation routes, and rescuing survivors and yet they are expected to perform at high proficiency levels should the situation arise.

The terms skill "loss", "retention", and "decay" have been used interchangedly in the literature. For purposes of clarity, we use the therm "skill loss" throughout the report.

In a recent review and meta-analysis of the skill loss literature, Arthur, Bennett, Stanush, and McNelly (in press) identified and quantified the effects of several factors on the loss of complex skills and knowledge after extended periods of non-use or non-practice. Specifically, they demonstrated that although the relationship between skill retention and the length of the non-practice interval is a negatively accelerated curvilinear relationship with the amount of skill loss ranging from about a *d* of -0.1 immediately after training (less than one day) to a *d* of -1.4 after more than 365 days of non-use, there is not one *single* decrement function curve but rather the decrement is specific to the task and situation. Factors that moderate this relationship include variables such as degree of overlearning; task characteristics such as procedural/control, physical/cognitive, natural/artificial, motor/verbal, and speed/accuracy; methods of testing; conditions of retrieval; lab/applied tasks; and evaluation criteria.

In addition to the variables listed above, Arthur et al. (1995) also sought to investigate the effects of instructional training methods and individual differences on skill loss. However, they were unable to do so because these variables have received very limited attention in the skill loss literature. The conceptual importance of each of these variables to skill loss is presented below.

Training Protocols and Instructional Strategies

One of the most pervasive weaknesses of the skill loss literature is the lack of attention given to the phenomena of skill acquisition (Arthur et al., 1995). Schmidt and Björk (1992), for example, criticize the educational and training literature for treating learning (i.e., skill acquisition), retention, and transfer as three separate phenomena which have been studied independently by different scientists using different methods in different laboratories. For instance, they show that manipulations which maximize performance during training can be detrimental in the long run. That is, those protocols which maximize skill acquisition may not necessarily lead to the best retention and transfer compared to other protocols which degrade speed of acquisition. Thus, these authors argue that acquisition, retention, and transfer

are really inseparable and need to be considered together when conducting research on skill acquisition. Therefore, in any investigation of long-term skill retention, the relationship between skill acquisition and skill retention is vitally important and needs to be taken into account. For example, a researcher or a training specialist can use all the "best" methods to facilitate retention by manipulating aspects of the retention interval and the retention testing situation, but if little or no skill is initially acquired during training, retention as a phenomenon, becomes a moot issue. Consequently, an objective of the present study was to evaluate the effectiveness and efficiency of two different training protocols (i.e., the Active-Interlocked Modeling [AIM]-dyad and an Individual-based protocol) in terms of complex skill acquisition, loss, and re-acquisition after a non-practice post-training interval.

The Active-Interlocked Modeling (AIM) Protocol. The *active interlocked modeling (AIM) protocol* is a dyadic training protocol developed by Shebilske, Regian, Arthur, and Jordan (1992). This protocol is defined as observational learning in the context of actively performing a task in harmony with a partner. The AIM-dyad protocol requires trainees to perform each half of a task alternately with a partner who performs the other half. The goal is for each trainee to learn both parts (in the present study, the pilot/gunner and mine/missile components of the Space Fortress task) by hands-on practice on alternate trials and to learn the connection between parts by modeling the actions and reactions of their partner. The increased efficiency of AIM results from the ability to train two people to achieve the same performance level as a single person with no increase in trainer time or machine cost (Shebilske et al., 1992).

Shebilske and associates have consistently achieved a 100% increase in training efficiency over a control individual-trainee based condition (e.g., Arthur, Strong, Jordan, Williamson, Shebilske, & Regian, 1995; Arthur, Young, Jordan, & Shebilske, 1996; Shebilske et al., 1992; Shebilske, Jordan, & Arthur, 1993; Shebilske, Jordan, Arthur, & Regian, 1993; Shebilske & Regian, 1992). Table 1 presents a meta-analysis of four published studies that used both the AIM-dyad and an individual-based protocol to train research participants on Space Fortress. Using the AIM-dyad protocol as the "experimental" group in the computation of the meta-analysis statistics, the results clearly demonstrate that there was practically no difference between the two groups in their performance on Space Fortress (δ = 0.01, SDδ = 0.00). Rerunning the meta-analysis in terms of the improvement in performance over baseline (i.e., pretest [baseline] = control and posttest [final session] = experimental) resulted in slightly better performance for the AIM-dyad protocol. The δ for the AIM-dyad and Individual-based protocols were = 3.23 (SDδ = 1.69) and 3.05 (SDδ = 0.84), respectively, resulting in a performance improvement difference δ of 0.18.

Table 1

Results of Meta-Analysis for the Performance Difference Between AIM-dayd and Individual-based Training Protocols on Space Fortress

Number of Data Points	Total Sample Size	Corrected Statistics: Mean d	δ	SDδ	% Variace due to Sampling Error	Min d	Max d	95% Confidence Interval: Lower Bound	Upper Bound
4	292	0.01	0.01	0.00	100.00	-0.06	0.04	0.01	0.01

[A]Study	AIM-DYAD Mean	SD	N	INDIVIDUAL-BASED Mean	SD	N	d
1	4267.02	902.02	23	4326.88	1251.84	24	-0.06
[B]2	1811.03	2034.65	46	1726.19	2017.69	52	0.04
3	4435.05	1027.12	20	4492.75	1238.15	20	-0.05
[B]4	1943.40	1933.44	30	1902.96	1863.09	77	0.02

NOTE: [A]1=Arthur et al.(1995); 2=Arthur et al.(1996); 3=Shebilske et al. (1992); 4=Prislin, Jordan, Worchel, Tschan Semmer, and Shebilske (1996). [B]Study included both female and male trainees; other studies had males only. Space Fortress means for each study are based on performance on the last session for the specified study. Mean *d* = sample-weighted mean *d*; δ=corrected mean *d*; SDδ=standard deviation of the corrected mean *d*. The confidence interval is used to assess the accuracy of the estimate of the mean effect size (δ). Specifically, it estimates the extent to which sampling error remains in the sample-weighted effect size.

Evaluating the effectiveness of training protocols in the context of skill loss is a logical extension of any research program or paradigm that seeks to assess the comparative effectiveness of specified training protocols. In other words, retention cannot be meaningfully separated from acquisition. Although Shebilske and associates have assessed the differential effectiveness of the AIM-dyad and other protocols, these studies have assessed performance on the training task in terms of relatively short time frames, that is, performance after 2-10 days. So although there is strong and consistent data about the effectiveness and efficiency gains associated with the AIM-dyad protocol (when compared to Individual-based protocols), there has to date been no assessment of the comparative effectiveness of these protocols in terms of long term skill loss and re-acquisition. Therefore, it is necessary to compare the effectiveness of these protocols using relatively more distal criteria. Although no specific hypotheses were formulated, one could ask whether the AIM-dyad and Individual-based training protocols differ in long term skill retention, and furthermore, whether they also differ in terms of training for skill *re-acquisition*. These long term tests are consonant with training in field and applied settings, particularly within the context of opportunity-to-perform (Ford, Quiñones, Sego, & Speer Sorra, 1992) and related issues.

Transfer of Skills/Training

In addition to the *amount* of skill acquisition, loss and re-acquisition, the comparative effectiveness of training protocols can also be assessed in terms of the facilitation or inhibition of the transfer of acquired skills. Transfer of training is a key criterion for evaluating the effectiveness of any formal training program (Kirkpatrick, 1987). Alliger, Bennett, and Tannenbaum (1995) draw a distinction between two definitions of transfer of training as used by Industrial/Organizational (I/O) psychologists and learning researchers. I/O psychologists have generally conceptualized transfer of training as representing the generalization of trained performance from the training environment to the work environment on a given task. Learning researchers, on the other hand, have often conceptualized transfer of training as involving the generalization of learning from one task to another, such that performance on Task B is facilitated (positive transfer) by training on Task A. Hence, the typical I/O conceptualization holds the task constant and the environment is modified; while the learning framework holds the environment constant and modifies the task. The present study used the latter framework.

Skill transfer is an important issue because it is not an uncommon occurrence for future performance tasks and sites to be different from those in which the individual was trained on. Thus, an individual may be trained to trouble shoot and repair F100 jet engines but may be later required to work on F200 jet engines which are substantially different from the F100. The importance of transfer is even more pronounced in the context of skill loss because a non-practice interval exists between the training (acquisition) and work (retention) environments and tasks.

In spite of the importance of this issue, it appears to have received almost no attention in the skill loss literature. The present study attempts to address this limitation in the current literature and extend Arthur, Bennett, Stanush, and McNelly's (in press) meta-analysis by focusing on transfer across tasks (the learning framework) instead of transfer across environments. In summary, a second objective of the present study was to engage in a preliminary investigation of the effect and role of skill loss in the *transfer* of acquired skills across both similar and dissimilar tasks. Specifically, we sought to comparatively assess the effectiveness of the AIM-dyad and Individual-based training protocols in terms of skill transfer to other tasks at three specific stages - during training (i.e., the acquisition phase), after training (the retention phase) and finally, during the re-acquisition phase.

Three transfer tasks were used in the present study. These were a keyboard version of Space Fortress and the computer games Asteroids (Logg, 1993), and Tempest (Theurer, 1993). These tasks were chosen because of procedural, skill-based, and strategic similarities and differences between them and the regular version of Space Fortress. In the absence of any empirical data about the extent to which these tasks, specifically, Asteroids and Tempest, were satisfactory transfer tasks, this study sought to assess the relationship between these two tasks and Space Fortress before trainees were trained; positive or negative relationships would indicate the extent to which these tasks can be considered to be reasonable positive or negative transfer tasks.

Although this research objective was primarily exploratory in nature, it was expected that trainees trained in the AIM-dyad protocol would be more successful at transferring their skill from Space Fortress to the keyboard version (and to Asteroids) compared to the trainees trained in the Individual-based protocol. This hypothesis is based on the rationale that trainees who learn the complex cognitive strategies underlying high performance of the Space Fortress task will successfully transfer these strategies to positive transfer tasks. Also, those trainees who rely on less complex cognitive strategies and more heavily on joystick control on the normal version will not perform as successfully on positive transfer tasks that do not require a joystick. Therefore, the aforementioned hypothesis is based on the assumption that trainees in the AIM-dyad protocol will rely more on complex cognitive strategies compared to trainees in the Individual-based protocol and that trainees in the Individual-based protocol will rely more on joystick control strategies compared to trainees in the AIM-dyad protocol. Based on this line of reasoning, AIM-dyad trainees were also expected to perform better compared to the Individual-based trainees on the transfer tasks at the retention and re-acquisition phases.

Finally, on the basis of its graphic interface, procedural, and operational rules, it was expected that Asteroids would serve as a positive transfer task. Specifically, it was anticipated that trainees who score better on Space Fortress would also perform better on Asteroids. Although the exact status of Tempest appeared to be more ambiguous, it was expected to serve and function as a negative transfer task primarily because its graphic interface, procedural, and operational rules are very different from those of Space Fortress.

Training to "Mastery"

A pervasive problem in the skill loss literature is the lack of consensus concerning the criteria used to determine the point at which skill acquisition should cease and the retention interval should begin (Arthur et al., in press). Many primary studies, for example, have trained individuals to one error-free trial (e.g., Goldberg, Drillings, & Dressel, 1981; Hagman, 1980a, 1980b; Schendel & Hagman, 1982) while other studies have used criteria such as a predetermined percentage of students correctly performing the task (e.g., Holgrem, Hilligoss, Swezey, & Enkins, 1979; Shields, Goldberg, & Dressel, 1979) as the point to end skill acquisition. Lastly, some studies have not specified a particular criterion that trainees must reach before skill acquisition was terminated; instead, trainees were required to complete a certain amount of training material (e.g., Adams & Hufford, 1962) or practice a certain task for a specified amount of time.

In addition to different types of criterion used to determine the termination of skill acquisition, differences in terminology is also a problem in the skill loss literature. One errorless trial, for example, has been labeled differently across studies (e.g., "proficiency," Hagman, 1980a; "minimal mastery," Farr, 1987; and "mastery," Hall, Ford, Whitten, & Plyant, 1983). The term "mastery" has also been used to refer to one error-free trial (Hall et al., 1983), to two error-free trials (Schendel & Hagman, 1982), and to three error-free trials (Goldberg et al., 1981).

In general, two methods of measuring skill acquisition have been used in the extant literature - namely (1) how much is trained in a specified amount of time and (2) how long it takes to train a certain amount of material. Although these criteria measure certain dimensions of performance, it cannot be assumed that they are interchangeable (cf. Adams & Hufford, 1962; Goldberg et al., 1981; Hagman, 1980a, 1980b; Holgrem et al., 1979; Schendel & Hagman, 1982; Shields et al., 1979). Furthermore, they do not in and of themselves guarantee or provide a hundred percent assurance that the task or skill has been mastered. However, for purely logistical reasons which arose from the use of a dyadic protocol, the present study used the former definition of "mastery", that is, to train trainees for a specified length of time.

Individual Differences

Research investigating the role of individual differences in skill loss has been very limited (Arthur et al., in press). Nevertheless, it has generally been argued and demonstrated that higher ability individuals, compared to lower ability individuals, retain more knowledge and skill over periods of non-use because they acquire more knowledge and skill in the same amount of time (Carron, 1971; Carron, & Marteniuk, 1970; Farr, 1987; Fox, Taylor, & Caylor, 1969; Grimsley, 1969; Purdy, & Lockhart, 1962; Schendel, Shields, & Katz, 1978; Vineberg, 1975). However, there is dissenting research which suggests that there is also a qualitative difference between high and lower ability individuals that may explain the enhanced skill retention exhibited by higher ability individuals. Farr (1987), for example, suggests that the differential loss rates observed between higher and lower ability individuals might be due to higher ability individuals using more effective strategies to acquire knowledge and skills. This is consistent with the findings of Hall et al. (1983) who required Navy sailors to complete two self-paced courses in basic electricity and electronics to a criterion of mastery. After a non-practice retention interval ranging from 18 to 34 days, Hall et al. (1983) found that higher ability sailors retained significantly more than lower ability sailors.

Individual Differences and Space Fortress

The primary task and dependent variable used in the present study was Space Fortress (Gopher, 1993; Shebilske et al., 1992). Space Fortress is a video game-like simulator. As a research tool, Space Fortress has an excellent record as a representative analogue of high-demand tasks (Donchin, 1989; Mane & Donchin, 1989). Described as "an experimental game which was designed to simulate a complex and dynamic aviation environment" (Gopher, 1993, p. 299), it has been used for both research and applied purposes. For instance, using Space Fortress as an analogue, researchers have investigated the effectiveness of various training protocols for complex perceptual-motor skill tasks (e.g., Frederiksen & White, 1989; Gopher, Weil, & Siegel, 1989; Shebilske et al., 1992).

From a more technical and applied perspective, performance on Space Fortress has also been demonstrated to transfer to actual flight performance (Gopher, Weil, & Bareket, 1994). Thus, this and other empirical research indicate that performance on this PC-based simulator is predictive of performance on similar real world high-demand complex

perceptual-motor skill tasks including those typically trained with automated instructional systems in current military and industrial applications (e.g., Donchin, 1989; Mane & Donchin, 1989; Gopher et al., 1994; Rabbitt, Banerji, & Szymanski, 1989).

A number of studies have investigated individual differences in the prediction of performance on Space Fortress. Rabbitt et al. (1989) looked at the relationship between Space Fortress performance and intelligence test scores. Intelligence predicted learning and performance on Space Fortress better than age and previous experience with video games. Foss, Fabiani, Mane, and Donchin (1989) looked at differences in individual performance on Space Fortress under unsupervised practice conditions. Although all trainees improved with time, differences were found between individuals in the initial capacity, learning rate, and strategies adopted. In a third study, Gopher et al. (1989) used an aiming test to divide trainees into two groups of high or low psychomotor ability. Trainees with high scores on the aiming test performed better on Space Fortress and were less influenced by training manipulations.

Arthur et al. (1995) investigated individual differences in selective attention as a predictor of performance on Space Fortress. Attention scores predicted performance before and after training. And, although training accounted for more variance, attention contributed significant incremental validity after training had been taken into account. Finally, Gottesfeld and Arthur (1994) examined the ability of two Big Five personality dimensions - Openness and Conscientiousness - to predict performance on Space Fortress. Their results indicated that although Openness correlated positively with performance across Space Fortress training trials, not all of these correlations were significant. Conscientiousness did not correlate significantly with performance across any of the Space Fortress trials. Finally, consistent with the preceding findings, neither personality variable contributed significant incremental validity over the prediction provided by either cognitive ability or training.

In summary, this literature suggests that there are identifiable individual difference variables that will predict complex task acquisition and performance, sometimes, over and beyond the effects of training. The present study extends this literature by investigating the effects of individual differences in the loss and re-acquisition of complex skills. Specifically, a third objective of the present study was to assess the ability of specified individual difference variables (i.e., cognitive ability, psychomotor ability, declarative knowledge, spatial working memory, spatial processing speed, and visual attention), to predict not only skill acquisition in original training, but also amount of skill loss and skill re-acquisition as well.

Stability of Ability-Performance Relationships as a Function of Training and Practice

An issue related to the ability of individual differences to predict not only skill acquisition in original training but also rate of skill loss and skill re-acquisition as well is the stability of ability-performance relationships over extensive time intervals (i.e., training and post-training). Although the prediction of individual differences in task performance has

had a relatively long history, there appears to be some recent disagreement about the stability of ability-performance relationships over time (e.g., Austin, Humphreys, & Hulin, 1989; Barrett, Alexander, & Doverspike, 1992; Barrett, Morris, & Alexander, 1993; Deadrick & Madigan, 1990; Hanges, Schneider, & Niles, 1990; Hulin, Henry, & Noon, 1990). One of the problems with previous literature investigating this issue has been the use of relatively simple information processing tasks which have been critiqued as being non-representative of real world complex tasks (Barrett et al., 1992). One dimension of representativeness is consistency of stimulus-response relationships. Some laboratory tasks, such as the pursuit rotor, have very consistent stimulus-response relationships. In contrast, many components of real world tasks, such as controlling air traffic or piloting a fighter jet, have inconsistent stimulus-response relationships. For example, the appropriate response to an enemy aircraft varies depending upon the situation. Ackerman (1992) argues that the relationship between individual difference measures and task performance changes as a function of training for tasks that have consistent information processing demands, but tends to remain the same for tasks with inconsistent stimulus-response relationships. Thus in support of the premise that components of Space Fortress contain inconsistent information processing demands, Arthur et al. (1995) demonstrated that attention-performance relationships remained stable over training as postulated by Ackerman's theory (e.g., Ackerman, 1992). Furthermore, in an extension of Ackerman's (1988, 1992) theory of dynamic ability determinants, Day, Arthur, and Shebilske (in press) demonstrated that the cognitive ability-performance relationship may increase over practice sessions in both the Aim-dyad and Individual-based protocol on Space Fortress.

However, all the preceding research has focused only on the acquisition phase with no attention paid to the retention and re-acquisition phases. Therefore, the current study examined ability-performance relationships not only over the acquisition phase but also over the loss and re-acquisition phases. On the basis of past research, it was expected that the ability-performance relationships observed in the acquisition phase will also be demonstrated in the re-acquisition phase since the stimulus-response relationships underlying the task are expected to be inconsistent in both of these two phases.

Research Questions

In summary, the general objective of the present study was to comparatively evaluate the effectiveness of the AIM-dyad and Individual-based protocols within the context of skill acquisition, retention, re-acquisition, and transfer. The present study extends the database on the short term effectiveness and efficiency of the AIM-dyad training protocol to include relatively more distal training effectiveness metrics that incorporate long periods of non-use or non-practice. Furthermore, this study permits statements concerning the effectiveness of the specified training protocols in training for skill re-acquisition. These objectives are important because the efficiency of the AIM-dyad protocol has inspired the pursuit of innovative dyadic protocols for training pilots and navigators in both military and non-military settings (Johnston, Regian, & Shebilske, 1994; Shebilske, Goettl, & Regian, in press).

The present study sought to answer the following research questions:

1. Will the amount and rate of skill acquisition be the same for the AIM-dyad and Individual-based training protocols?

2. Will the AIM-dyad and Individual-based training protocols result in different amounts of skill loss?

3. Will the amount and rate of skill re-acquisition be the same for the AIM-dyad and Individual-based training protocols?

4. Are there any differences between the AIM-dyad and Individual-based protocols in terms of performance on the transfer tasks?

5. What is the nature of the ability (i.e., cognitive ability, declarative knowledge, psychomotor ability, spatial processing speed, spatial working memory, and visual attention) and performance relationships over the acquisition, loss, and re-acquisition phases of task performance? Will these relationships be influenced by the training condition to which trainees were assigned?

METHOD

Participants

The final study sample consisted of 89 paid volunteers from a large southwestern university and its community who were recruited by an advertisement in the school newspaper and posted notices around campus. Trainees were paid $75 for their participation. Because of a number of logistical reasons (e.g., limited laboratory space and computers) trainees were run in five sequential groups. Trainees competed within their group for three bonuses - $100, $60, and $40 - which were awarded to the three trainees with the highest Space Fortress test scores over the 15 days of training. Thus a total of 15 bonuses were awarded. All trainees were male and right-handed. The mean age of the sample was 20 years, 6 months (*SD* = 3 years, 2 months; Min = 15 years; Max = 31 years).

Measures - PREDICTORS

As part of a larger project, trainees completed several measures only a subset of which are used in the present study. These measures are presented below. Descriptions for all the other measures can be found in Appendix A.

Aiming Task. The aiming task was used as a measure of psychomotor ability (Gopher et al., 1989). It is a short task which tests the speed and accuracy with which the participant can aim at and hit a target. Trainees used a joystick to control a spaceship located centrally on a computer monitor. Every time a mine appeared on the screen, the participant's task was to shoot and destroy the mine as quickly as possible. If the mine was hit it disappeared immediately, and if missed, it disappeared after a few seconds. Another mine then appeared a few seconds later. Scores were earned for each mine hit. Other points were either added to or subtracted from a speed score depending on whether or not the participant hit the mine and if so, how long it took to hit the mine since its appearance on the screen. The mine and speed scores were summed to create a total score for the aiming task. To maximize this total score, the participant had to respond as quickly and accurately as possible to the mines. The mean score of three, two-minute aiming tasks was used as a measure of trainees' psychomotor ability. Using each of the three administrations as an item, an internal consistency estimate of 0.84 was obtained for the aiming task in the present study.

Computer-Administered Visual Attention Test (CA-VAT). The CA-VAT (Arthur, 1991; Arthur et al., 1995; Arthur, Strong, & Williamson, 1994) is an IBM-based PC administered and scored test of visual attention. The general design of the test, constructed as an approximate visual counterpart to the Auditory Selective Attention Test (ASAT; Gopher & Kahneman, 1971; Mihal & Barrett, 1976), is based on protocol developed for the Visual Selective Attention Test (VSAT) by Avolio, Alexander, Barrett, & Sterns, (1981). The stimuli in the CA-VAT are pairs of numbers and letters that appear on a computer monitor. A given pair of characters consists of either two numbers, a number and a letter, or two letters. Cue words preceding each item signal the appropriate response sequence. There are 12 items and a test taker's score is the number of errors made.

Because test takers usually require considerable practice with the CA-VAT before they are familiar with the instructions and required keystroke responses (Arthur, 1991; Arthur et al., 1994; Strong, 1992), two forms of the CA-VAT are administered with the first being used as a practice session - although the trainees are not informed of this (Arthur et al., 1994). Arthur et al. (1994) report moderate convergent validity for the CA-VAT and ASAT ($r=0.25$) and internal consistencies of 0.93 to 0.98. A test-retest reliability of 0.83 has also been reported for the CA-VAT (Strong, 1992). The CA-VAT has also been found to correlate 0.05, 0.07, and 0.10 with computer attitudes, computer familiarity, and computer intimidation, respectively, and 0.32 to 0.37 with general cognitive ability (Arthur, 1991). Additional predictive validity for the CA-VAT has been demonstrated by Arthur et al. (1994). An internal consistency estimate of 0.88 was obtained for the present study.

Declarative Knowledge Test. A 30-item, paper-and-pencil, two to four alternative, multiple-choice, content valid test was constructed to assess trainees' knowledge of the Space Fortress instructions, procedural rules, and information that was presented to them both via video and written text. Trainees had seven minutes to complete the test. A participant's score was the total number of questions answered correctly. As indicated in Table 3, this test was administered on three separate occasions. The mean intercorrelation between the three administrations was 0.59. The Declarative Knowledge Test items are presented in Appendix B.

Figure Matrices Test (g; Kyllonen, Christal, Woltz, Shute, Tirre, & Chaiken, 1990). The Figure Matrices test is part of the Cognitive Abilities Measurement (CAM) test battery (version 4.0) developed at the Brooks Air Force Base. It is a computer-administered test of cognitive ability (*g*) analogous to the paper-and-pencil Raven Advanced Progressive Matrices test (Raven, Court, & Raven, 1993). The test consists of nine, progressively difficult items that require the trainees to choose which of eight options best completes a pattern series presented across three rows of designs. The last design in the final row (third column, third row) is blank; thus the participant's task is to choose the piece that best completes the overall matrix. Trainees have 65 seconds to answer each item with a 5 second warning after 60 seconds. The average time to completion is approximately five minutes. An algorithm that took into account both the answer (right/wrong) and the response time to each item was used to generate the test score. Specifically, an item score was computed as [A x B] + [B - D] where A = answer (right answer = +1; wrong answer=0), B = maximum response time (i.e., 65 seconds), and D = response time. The test score was the sum of the item scores. Split-half and test-retest reliabilities of 0.68 and 0.59 respectively, have been reported (Kyllonen et al., 1990). Gottesfeld and Arthur (1994) report a coefficient alpha of .62. These psychometric data were all based on a right/wrong answer scoring scheme. An internal consistency estimate of 0.48 was obtained for the present study.

Spatial Processing Speed Test (SPST; Kyllonen et al., 1990). The Spatial Processing Speed Test is also part of the CAM test battery. It is a 12-item computer-administered test of (spatial) processing speed. In this test, trainees must decide, as quickly as possible, whether presented figure combinations match the simple sequence formula

specified by figure statements initially presented. Figures consist of blocks divided by a diagonal line, and colored pink with black, or blue with black (e.g., one side of the diagonal is black and the other is pink). The direction of the diagonal changes positions, allowing for different combinations (e.g., a diagonal going from top-left to bottom-right may cause pink to be on the top and black to be on the bottom; a diagonal going from top-right to bottom-left may cause the black to be on top and the pink on the bottom). For each item, two blocks of the same color (i.e., either pink/black or blue/black) appear with an arrow. The arrow describes the sequence in which these two blocks should appear (e.g., one on top of the other). The arrow sometimes has a slash through it, which is interpreted as meaning "not" (e.g., Block 1 will not appear below Block 2).

Shortly after the figure statement is presented, a set of two blocks is shown in the middle of the screen. Trainees must decide as quickly as possible whether this combination matches what is described in the initial statement. If they match, "L" (for "like") is the correct response; if they do not match, "D" (for "different") is the correct response. Upon entering the correct response music is sounded. Incorrect responses are followed by a buzzer. The next item is then presented, preceded by three warning asterisks.

Trainees have 15 seconds to answer each item, and the average time to completion is approximately five minutes. An algorithm that took into account both the answer (right/wrong) and the response time to each item was used to generate the test score. This algorithm was similar to that used for the Figure Matrices Test. The test score was the sum of the item scores. Split-half and test-retest reliabilities of 0.70 and 0.42 respectively, have been reported (Kyllonen et al., 1990). These psychometric data were all based on a right/wrong answer scoring scheme. An internal consistency estimate of 0.76 was obtained for the present study.

Spatial Working Memory Test (SWMT; Kyllonen et al., 1990). The Spatial Working Memory Test is also part of the CAM test battery. It is a 24-item computer-administered test of (spatial) working memory. In this test, trainees are required to relate what is described in three pictorial statements to a sequence of four block figures. Figures consist of blocks divided by a diagonal line, and colored pink with black, or blue with black (e.g., one side of the diagonal is black and the other is pink). The direction of the diagonal changes positions, allowing for different combinations (e.g., a diagonal going from top-left to bottom-right may cause pink to be on the top and black to be on the bottom; a diagonal going from top-right to bottom-left may cause the black to be on top and the pink on the bottom). For each statement, two blocks of the same color (i.e., either pink/black or blue/black) appear with an arrow. The arrow describes the sequence in which these two blocks should appear (e.g., one on top of the other). The arrow sometimes has a slash through it, which is interpreted as meaning "not" (e.g., Block 1 will not appear below Block 2). The third statement merely displays solid pink and solid blue blocks, describing the sequence of the pink and blue blocks (e.g., pink will not appear before blue).

The pictorial statements appear one at a time at the top of the screen. Trainees must determine the sequence of blocks as these statements appear. After the final statement, eight numbered alternatives appear on the screen with a timer. These alternatives represent the possible combinations using the presented statements. Using the 1-8 number keys at the top of the keyboard, the participant's task is to type in the number corresponding to the correct sequence before the 15 seconds on the timer runs out. Correct responses are followed by music. Incorrect responses are followed by a buzzer, and the three statements are then displayed to show how the participant's response is incorrect. Next, three asterisks appear to warn the participant that the next item is about to be presented.

Trainees have 15 seconds to answer each item, and the average time to completion is approximately 17 minutes. Again, an algorithm that took into account both the answer (right/wrong) and the response time to each item was used to generate the test score. This algorithm was similar to that used for the Figure Matrices Test. The test score was the sum of the item scores. Split-half and test-retest reliabilities of 0.86 and 0.69 respectively, have been reported (Kyllonen et al., 1990). These psychometric data were all based on a right/wrong answer scoring scheme. An internal consistency estimate of 0.71 was obtained for the present study.

Video Game Experience Questionnaires A and B. Two short questionnaires (5 and 3 items, respectively) were developed to collect data on the extent to which trainees had played the two transfer tasks (i.e., Asteroids and Tempest) before signing up to participate in the study (Questionnaire A), and also during the eight week non-practice interval (questionnaire B). Both measures also asked trainees to rate their ability levels on the transfer tasks/games on a five-point Likert scale (1 = novice, 3 = average, 5 = expert). Questionnaire A was administered during the screening session and Questionnaire B was administered during Session 10, the first session after the retention interval. This data was collected to test for differences between the training protocols on Asteroids and Tempest before the commencement of training and also after the 8-week non-practice interval.

CRITERIA

Space Fortress. Space Fortress (Gopher, 1993; Shebilske et al., 1992) is a complex perceptual-motor skill task specifically designed to be a representative analogue of complex high-demand tasks (Mane & Donchin, 1989) and has also been described as "an experimental game which was designed to simulate a complex and dynamic aviation environment" (Gopher, 1993, p. 299). Space Fortress requires the following equipment: an IBM AT compatible with an 80386 processor or higher a (math co-processor is required for a 80386 or 80486SX), a VGA monitor, a joystick and a three-button mouse. In this task, a fortress occupied the center of the monitor screen in the middle of the smaller of two concentric hexagons. A drawing of the Space Fortress screen is presented in Figure 1. The participant controlled a space ship's flight path with the joystick and shot missiles with a trigger button on the joystick. The three-button mouse was used to perform functions related to the mines and bonuses. The right-button was used to "prime" foe mines (two button presses with an interpress interval of 250-400 msec) before they could be destroyed.

The left-button was used to select a points bonus and the middle-button was used to select a missile bonus when the second of a pair of "$" symbols appeared.

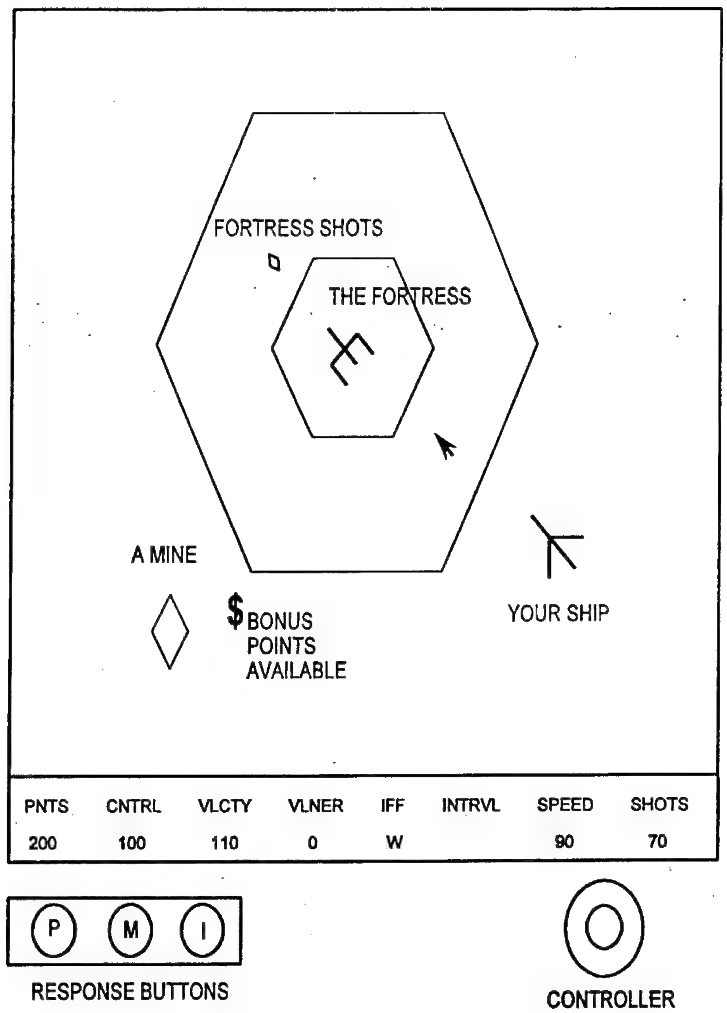

Figure 1. Space Fortress Game Screen, Left-Handed Response Buttons, and Right-Handed Joystick

Mines appeared every four sec and stayed on the screen for 10 sec unless they were hit by a missile from the ship, collided with the ship, or until the ship was hit by the fortress. A single fortress hit damages the ship with a concomitant loss of 50 points; four hits destroy the ship with a concomitant 100-point loss. While mines were on the screen, the ship's missiles were ineffective against the fortress.

Before each three minute game, trainees memorized three computer generated letters which identified foe mines. A letter always appeared in the identify-friend-or-foe (IFF) indicator when a mine appeared on the screen. If the letter indicated a foe mine, trainees had to press the IFF button twice with an interpress interval of 250-400 msec. The information panel displayed this interval. If it was incorrect, they could try again. Before they got the correct interval, the ship's missiles were ineffective against the mine. After they got the correct interval, they could destroy the mine with one missile hit. The mine continuously chased the ship attempting to destroy it by colliding with it. If the IFF letter represented a friendly mine, trainees were to avoid hitting the IFF button. If they did hit the IFF button, they turned the friend mine into a deadly enemy which behaved like a foe mine, but was worse because it could not be destroyed. If the IFF button was not pushed, the friendly mine pursued the ship waiting to be energized by a missile, which would score points and increase the fortress vulnerability counter by one. Even a friendly mine could destroy the ship if it collided with it before an energizing missile was delivered.

When mines were not on the screen, the ship's missiles could damage and eventually destroy the fortress. Each of the first 10 missile hits on the fortress increased its vulnerability. After 10 hits, it could be destroyed by a double shot, which had to have an interpress interval of 250 msec or less. If a double shot hit the fortress before the tenth hit, the vulnerability was reset to zero.

The ship started with 100 missiles, which was its main supply. It could fire more missiles after the main supply was depleted but it cost three points per missile to do so. The main missile supply could be replenished during bonus intervals, which were indicated by the second of two consecutive "$" symbols appearing below the fortress. Symbols changed every four sec and consisted of other symbols in addition to the "$" symbol. The "$" symbol always appeared in consecutive sets of two. If the participant pressed a bonus button during the first of the pair, no bonus was delivered, and the bonus buttons were deactivated during the second "$" symbol, the bonus interval. However, if the participant waited until the second "$" symbol appeared, they could press the missile bonus button to receive extra missiles. If they had more than 50 missiles remaining, their total would be restored to 100. If they had less than 50 missiles remaining, 50 more would be added. Alternatively, the participant could press the points bonus button to obtain 100 extra points during the bonus interval.

The spaceship flew in frictionless space so a thrust in one direction would move the ship at a constant velocity in that direction until another thrust was applied. Thrusts in the same direction accelerated the ship, thrusts in the opposite

direction slowed the ship, and thrusts in other directions changed the ship's course. Pushing the joystick forward applied thrust, moving the joystick left or right rotated the ship, and pulling back on the joystick did nothing. If the ship left the edge of the screen, it reappeared on the opposite side of the screen.

Part of the total score was determined by points, which were awarded as follows: 4 for hitting the fortress, 100 for destroying it, 20 for energizing a friendly mine, 30 for destroying a foe mine, and 100 for selecting the points bonus when it was available. Points were lost as follows: -3 for firing after depleting the ship's missile supply, -50 when the ship was damaged by fortress or mine, and -100 when the ship was destroyed.

Ship control also contributed to the total score. Control points accumulated at a rate of 7 points per second when the ship was between the small and large hexagons surrounding the fortress. This rate of points gain was halved when the ship was outside the large hexagon. Control points could also be lost as follows: -35 for going off the computer screen and -5 for entering the inner hexagon.

Ship velocity also affected the total score. Velocity points were accumulated at a rate of 7 points per second as long as the ship remained below a critical velocity; no velocity points were gained when the velocity exceeded the critical velocity.

Finally, the speed with which mines were handled contributed to the total score as well. These speed points could range from -50 to +100 for friend mines and from -50 to +150 for foe mines. In the current study, the *total score* was used as the Space Fortress performance score. Trainees were always aware of their performance because their scores were presented "on-line" at the bottom of the computer screen. Thus at the end of each game and session, trainees knew exactly how well or poorly they had performed.

Trainees in both training conditions were informed of four strategies which other researchers (e.g., Frederiksen & White, 1989) have determined to be optimal and have been used in subsequent studies (e.g., Arthur et al., 1995; Shebilske et al., 1992). Trainees were instructed to slowly circumnavigate the fortress while remaining within the two hexagons. They were also to select point bonuses instead of missile bonuses unless the missile supply was below 50. Trainees were instructed not to chase mines when they appeared. They were to stay on course around the fortress and let the mines come close to the ship, then turn and fire at them. Finally, if or when the IFF button was pressed for a friendly mine, trainees were instructed to let the mine destroy the ship; they were not to waste time running from it. These strategies were conveyed to the trainees because most applied training situations include optimal strategies that have been developed by experienced operators. In contrast to many research tools, Space Fortress is complex enough and has been utilized enough to have well documented optimal strategies (e.g., Frederiksen & White, 1989).

Ship Control Strategy. A rating form was designed to assist proctors in rating trainees' ship control strategy. Four variables were coded - ship maneuvering strategy, ship speed (1 = slow, 5 = fast), joystick manipulation (1 = minimal, 5 = maximal), and course deviation due to mines (1 = low, 5 = high). Ship maneuvering strategy was rated as being one of seven categories, namely, three types of circling strategy (clockwise, counter clockwise, and random), three types of wrapping (vertical, horizontal, and diagonal), and an unspecific category (i.e., any strategy that did not fit any of the preceding categories; for these, the proctor was also expected to provide additional descriptive information on any strategy that was coded as such). A description of these variables and the reason for the ordinal rating used to operationalize them is presented below.

Ship-Maneuvering Strategies: Clockwise circling is the most effective strategy for maneuvering the ship around the Fortress. The ship starts out in a vertical orientation pointed at the top of the screen. As soon as the game begins, the Fortress starts firing at the ship, so the fastest way to avoid being hit is to fly the ship up, which, when changed to circling, results in clockwise circling. Counterclockwise circling is as effective as clockwise circling except when the ship is destroyed and must start over in front of the Fortress. When this occurs, the participant must rotate the ship before thrusting to circle counterclockwise that results in the Fortress being able to shoot the ship at least once. Random circling is the least effective of the three circling strategies because it relies on excessive joystick manipulation and too much time is spent with course corrections that limits the amount of time available for gaining points. Nevertheless, these three circling strategies are better than the wrapping or unspecified strategies described by Frederiksen and White (1989).

The vertical wrapping strategy is, by far, the easiest, and in terms of energy expenditure, the most efficient strategy for maneuvering the ship during the task. All the participant must do is apply thrust to the ship when the game starts or after the ship has been destroyed and starts over. This thrust results in vertical wrapping leaving the participant only to rotate the ship and fire at the Fortress and mines as the ship moves by them. However, this strategy is not effective in terms of trainees' performance scores because of the points that are lost due to wrapping the screen and the time and opportunities lost waiting for the ship to move into the correct position for shooting the targets. Similarly, horizontal wrapping is not as effective as vertical wrapping because of the maneuvering effort required to get the ship into a horizontal flight pattern. Diagonal wrapping is the least effective of the three wrapping strategies because of the extra effort that must be used with the joystick to maneuver the ship from colliding with obstacles and the inability to determine where the ship will reappear when it wraps the screen. Unspecified ship-maneuvering strategies are those which do not clearly fall into one of the above strategies, or is some combination of some or all of them. These strategies tend to require large amounts of joystick manipulation in order to control the ship which leaves little attentional resources and time for other aspects of the task.

Therefore, based on Frederiksen and White's (1989) component analysis of ship-maneuvering strategies, the strategies used in Space Fortress are (from most effective to least effective): 1) clockwise circling; 2) counter-clockwise circling; 3) random circling; 4) vertical wrapping; 5) horizontal wrapping; 6) diagonal wrapping; and 7) unspecified ship maneuvering strategy.

Ship Speed: One goal of flying the ship in the Space Fortress environment is to maintain a very slow ship speed. There are several reasons why this is important. The most important is the fact that a slowly moving ship is easier to control and aim than a fast moving ship. Specifically, by flying the ship at a slow speed, the participant is better able to maintain a course within the hexagons. Also, the ship can be more easily aimed at the Fortress and mines without fear of wrapping the screen. In addition, by moving the ship slowly, more attentional resources can be diverted to other activities because the slowly moving ship does not require as much attention to control it as a fast moving ship would.

Joystick Manipulation: Joystick manipulation is very important in flying the ship in the Space Fortress task. An effective strategy for joystick manipulation is to only move the joystick left or right when counter-clockwise or clockwise movement is desired; or forward to generate thrust. It is not effective to move the joystick in any diagonal direction or backward as these movements do not produce optimal results. Unintentional thrusts can seriously undermine effective ship control. For example, whenever the joystick is pushed forward the ship accelerates in the direction it is pointing. One way to avoid this is to avoid pushing the joystick forward unless ship thrust is desired. Gripping the joystick lightly and moving it in small, discrete increments results in better ship control dynamics than gripping the joystick tightly and moving it in large, continuous movements. Therefore, minimal joystick manipulation is essential for effective ship control.

Course Deviation due to Mines: One of the optimal strategies outlined by Frederiksen and White (1989) suggests how to control the ship when different mine scenarios are occurring. They found that the most effective strategy was to always maintain course and turn to shoot the mine when it approached close enough. An example of a sub-optimal strategy is to follow the mine around the screen while trying to destroy it. Course should be maintained even in the event of an erroneous IFF button press that causes a friend mine to be transformed into a foe mine that cannot be destroyed. In this case, it is more advantageous to let the mine hit the ship than to try and evade it. Trying to evade the mine is a sub-optimal strategy because by avoiding the mine until it disappears, the ship loses valuable time that could be spent shooting at the Fortress; points are also lost for leaving the hexagons and wrapping the screen. On the other hand, by letting the mine hit the ship, only the points for a ship hit are lost as well as a minimal amount of time which, in turn, allows the participant to reinstate the attack against the Fortress faster than would be the case if they tried evade the mine.

For each test session, proctors rated trainees on each of the four strategy components described above. An overall strategy score was the sum of the four components. These scores were reversed for the analyses such that a higher score represented higher levels of ship control.

Proctor Strategy Rating Training. Because proctors varied across trainees and sessions, an extensive training program was employed to minimize and control for potential rater effects. The first step was to inform proctors about the purpose and importance of collecting strategy ratings in the study. Next, they were introduced to the Strategy Rating Form and each dimension was described and defined in depth. An on-screen demonstration was then performed to give the proctors an appropriate frame of reference for each level of each dimension (i.e., horizontal wrapping vs. vertical wrapping; low joystick manipulation vs. high joystick manipulation, etc.). To practice rating an actual session, the proctors used the Strategy Rating Form to rate an experienced player for three sessions. After each of these three sessions, ratings were discussed and any questions and deviations were addressed. When there were no more questions from the proctors, each proctor individually rated an experienced player for five consecutive sessions; this player adopted specific pre-determined strategies for each of these sessions. These data were subsequently analyzed to obtain measures of inter-rater reliability and accuracy. The outcomes of these analyses are presented in the results section.

Asteroids©. Asteroids (Logg, 1993) requires an IBM PC/AT computer with a standard keyboard. Trainees control a spaceship and must shoot and maneuver their way through an asteroid belt. The goal is to shoot missiles at and destroy the asteroids before they collide with the ship. The game starts with a spaceship in the center of the computer screen surrounded by a group of large, moving asteroids. These large asteroids break into two medium-sized asteroids when hit by a missile. These medium-sized asteroids, in turn, become two small asteroids when shot. Large or small flying saucers appear on the screen in a random and unpredictable fashion and actively shoot at the spaceship. A flying saucer will destroy the ship unless it is shot first or collides with an asteroid. However, if neither the trainees' spaceship nor the enemy saucer are destroyed, the enemy saucer will leave the screen after an interval of 8 seconds. Before each game, trainees are provided with a "Quick Help Screen" which explains the basics of how to play the game. The participant is required to press the spacebar to fire a missile. The left-arrow key or the right-arrow key are pressed to turn the spaceship left or right, respectively. Thrust is applied to the spaceship by pressing the up-arrow key. The more times or the longer the up-arrow key is depressed or held down, the more thrust is applied to the spaceship. The spaceship operates in a frictionless environment and therefore continues to move when the participant stops pressing the up-arrow key. To stop the spaceship, participant must press the left- or right-arrow key until the ship is pointed in the opposite direction followed by the application of thrust by pressing the up-arrow key. This procedure is known as "reverse thrust". Thus the procedures for flying and managing the spaceship are very similar to those for Space Fortress.

The participant is also capable of escaping into hyperspace by pressing the shift key which causes the ship to reappear in another part of the asteroid belt. However, there is the chance that the spaceship will be destroyed upon reentry by colliding with an asteroid. Participant initially receive three spaceships at the start of each game and receive an extra ship for each 10,000 points accumulated. Shooting a large asteroid yields 20 points, a medium asteroid 50 points, and a small asteroid 100 points. The large flying saucer is worth 200 points and the small flying saucer 1000 points.

Trainees were instructed to score as many points as they could in six minutes. A participant's score was, therefore, the sum of the points obtained on the total number of games played in six minutes.

Space Fortress Keyboard Version. The keyboard version of Space Fortress is designed to be a transfer task of the normal version of Space Fortress. The keyboard version is identical in all respects to the normal version except that joystick control is replaced by keyboard control. The mouse functions remain the same. On a 101-key PC/AT keyboard the participant uses the cursor positioning arrow keys located to the left of the numeric keypad to control the movement and firing of the ship. The up-arrow key applies thrust to the ship in the direction it is pointing. The right- and left-arrow keys rotate the ship clockwise and counter-clockwise, respectively. The down-arrow key fires the ship's missiles.

Tempest©. Tempest (Theurer, 1993) requires an IBM PC/AT computer with a standard keyboard and mouse. Trainees control a ship called a "Blaster" that travels around the rim of what is known as a universe. The goal is to shoot and destroy "aliens" as they attempt to travel up corridors or tubes to the rim where the "Blaster" is situated. Before each game, trainees are provided with a "Quick Help Screen" which explains the basics of how to play the game. Trainees are required to move the mouse to align the "Blaster" with a corridor having an "alien" in it and click the left mouse button to fire a missile at the "alien". The participant also has at their disposal a "SuperZapper" which destroys all "aliens" on the screen. However, only two "SuperZapper" firings are permitted at each level. The "SuperZapper" is initiated by pressing the spacebar. Trainees initially receive three "Blasters" at the start of each game and receive an extra one for each 20,000 points accumulated. Upon destruction of all "aliens" moving up the corridors, the participant's "Blaster" travels down a tube into hyperspace to a new level. However, the participant must position their "Blaster" over a corridor that is free from spikes so as not to be destroyed as it moves to the next level.

There are many different types of "aliens" and each type of "alien" is worth a different number of points. Shooting a "spiker" yields 50 points, a "tanker" 100 points, and a "flipper" 150 points. Trainees also receive 200 points for shooting a "pulsar" and 250, 500, or 750 points for shooting a "fuseball" (more points are scored the closer it is to the "Blaster" when shot).

As with Asteroids, trainees were instructed to score as many points as they could in six minutes. A participant's score was, therefore, the sum of the points obtained on the total number of games played in six minutes.

Training Protocols

Trainees were randomly assigned to one of two training conditions - either the AIM-dyad or Individual-based protocol after the screening session described below. The AIM-dyad protocol was characterized by the presence of active-interlocked modeling, social contact, distributed and variable hands-on practice, and four hands-on practice trials per practice-trial session (Arthur et al., 1995). The Individual-based condition did not have the above factors, but trainees had eight hands-on practice trials per practice-trial session.

[Level 3] AIM-dyad Condition. The final sample size for the AIM-dyad condition was 40. The trainees in the AIM-dyad condition practiced the Space Fortress task with a partner for eight, three-minute games. One participant controlled all functions related to the joystick and trigger (pilot/gunner) while the other controlled all functions related to the response buttons (mine/missile manager). AIM-dyad trainees alternated roles after every practice game so they controlled each half four times. Partners were strongly encouraged to communicate about the task. For example, the mine/missile manager was advised to tell the pilot/gunner whether the mines were friend or foe and when the ship's missiles would be effective against the mines. Although the data is not used in this report, vocalizations during practice sessions were unobtrusively recorded by means of hidden microphones and tape recorders to empirically assess the amount of communication that actually took place between partners. Trainees in this condition were assigned to the same partner throughout training.

Individual-based Condition. The final sample size for the Individual-based condition was 49. Trainees in the Individual-based condition performed eight, three-minute practice games alone while operating both the joystick and response buttons simultaneously. Like the AIM-dyad conditions, trainees' vocalizations were also recorded during practice sessions.

Procedure

Screening Session. When trainees were recruited, they were informed that they would be performing several video game-like computer tasks and completing measures of individual differences and characteristics. The screening session entailed the completion of a battery of tests and measures, most of which were computer-administered. Before starting the screening session, trainees first signed informed consents and contract-for-pay forms. They then completed the battery of tests listed in Table 2. This screening session lasted about two hours.

Table 2

Design and Data Collection Procedures

DAY	TASK
Screening	Video Game Experience Questionnaire - Addendum A Aiming Task Video Game Experience Questionnaire ^Computer-Administered Visual Attention Test - 1st administration Goldberg's 100 Unipolar Markers Computer-Administered Visual Attention Test - 2nd administration Marlowe/Crowne Social Desirability Scale Spatial Processing Speed Test (Two-Term Ordering) Figure Matrices (*g*) NEO Five-Factor Inventory Spatial Working Memory Test (Four-Term Ordering) Self-Monitoring Scale Computer Attitude Scale Driving Behavior Questionnaire Interaction Anxiousness Scale
Monday	Asteroids [6 min.] Tempest [6 min.] Space Fortress Instructions [Video] Space Fortress Session 0 [4 games, and Strategy Rating] Space Fortress Summary of Instructions [Video] Declarative Knowledge Test
Tuesday	***Summary of Space Fortress Instructions placed at all stations*** Confidence/Alertness Questionnaire Space Fortress Session 1 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Wednesday	Space Fortress Session 2 [8 practice games, 2 test games, and Strategy Rating]
Thursday	Confidence/Alertness Questionnaire Space Fortress Session 3 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Friday	Space Fortress Session 4 [8 practice games, 2 test games, and Strategy Rating]
Monday	Confidence/Alertness Questionnaire Space Fortress Session 5 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Tuesday	Space Fortress Session 6 [8 practice games, 2 test games, and Strategy Rating]
Wednesday	Confidence/Alertness Questionnaire Space Fortress Session 7 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Thursday	Space Fortress Session 8 [8 practice games, 2 test games, and Strategy Rating]
Friday	***Summary of Space Fortress Instructions removed from all stations*** Confidence/Alertness Questionnaire Space Fortress Session 9 [8 practice games, 2 test games, and Strategy Rating] Declarative Knowledge Test Unprepared Simple Reaction Time Test Asteroids [6 min.] Tempest [6 min.]
	EIGHT [8] WEEK NON-PRACTICE INTERVAL
Monday	***Summary of Space Fortress Instructions removed from all stations*** Declarative Knowledge Test Video Game Experience Questionnaire - Addendum B Space Fortress Session 10 [2 test games, and Strategy Rating] Space Fortress Session 11 [2 test games, and Strategy Rating] Asteroids [6 min.] Space Fortress Session 12 [Keyboard Version - 2 test games] Tempest [6 min.] Space Fortress Session 13 [2 test games, and Strategy Rating]
Tuesday	***Summary of Space Fortress Instructions placed at all stations*** Confidence/Alertness Questionnaire Space Fortress Session 14 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Wednesday	Space Fortress Session 15 [8 practice games, 2 test games, and Strategy Rating]
Thursday	Confidence/Alertness Questionnaire Space Fortress Session 16 [8 practice games, 2 test games, and Strategy Rating] Unprepared Simple Reaction Time Test
Friday	***Summary of Space Fortress Instructions removed from all stations*** Space Fortress Session 17 [8 practice games, 2 test games, and Strategy Rating] Asteroids [6 min.] Space Fortress Session 18 [Keyboard Version - 2 test games] Tempest [6 min.] Space Fortress Session 19 [2 test games, and Strategy Rating]

The aiming task (Arthur et al., 1995; Mane & Donchin, 1989; Shebilske et al., 1992), which was part of this battery, was later used to screen trainees. This task was also used as a measure of psychomotor ability and has been previously described. However, for screening purposes, the highest score of three, two-minute aiming tasks was used as the criterion to eliminate or retain trainees (Mane & Donchin, 1989). Specifically, potential trainees were not permitted to participate if they failed to obtain a minimum aiming score of 780 points or if they reported playing more than 20 hours of video games per week (Shebilske et al., 1992). This screening was used to reduce error variance and to make the trainees more representative of those in operational training centers. Two individuals out of the total number who were recruited failed the aiming task and nine reported playing more than 20 hours of video games per week.

Five trainees whose visual attention scores were on the average 90% of the total number of possible errors on the CA-VAT were dropped from the data set. An additional 77 trainees were dropped from the data set due to incomplete data (e.g., they did not complete or participate in all training sessions; parts of their Space Fortress and/or other data were unavailable due to computer malfunctions, etc.). Thus out of an initial 182 trainees who were recruited to participate in the study, the final sample size was 89

Training and Test Sessions. Each participant was paid $75 to participate in 15 one-hour sessions held on Monday through Friday on two consecutive weeks followed by an eight-week non-practice interval, and then a final Monday-Friday block of sessions. Trainees were paid $1 per session if they terminated their participation before the end of the last (15th) session or if they (or their partner, in the case of dyads) missed a session which could not be rescheduled that same day thus calling for their removal from the study.

Trainees were informed that they were competing within their group for three bonuses - $100, $60, and $40 - which were awarded to the three trainees with the highest Space Fortress test scores over the 15 days of training. Trainees were informed of the bonuses before the study, but were not told how other trainees were doing.

On Day 1 (the first Monday), trainees who passed the screening were given standardized Space Fortress instructions, and participated in a procedural rules training session. This training was presented via video. Trainees were also given a copy of the video-taped script and were instructed to follow and read along while they watched the tape. After this training session, trainees completed a Space Fortress baseline session (Session 0) which consisted of four 3-minute games. They were next presented a summary of the Space Fortress instructions via video-tape; as with the first/full instructions training, they were also given a copy of the video-taped script and were instructed to follow and read along while they watched the tape.

Trainees were then administered the Declarative Knowledge Test; trainees had not been previously informed that they would be taking this test. After this, they were randomly assigned to the two training protocols.

The specific procedures and protocols used to run the trainees are presented in Table 2. Each Space Fortress session consisted of 8 practice games and two test games. The Space Fortress test games were always performed in the same manner across protocols with each participant performing the whole task alone. The practice games were identical to the test games for the Individual-based protocol, but were different from the test games in the AIM-dyad condition because dyadic trainees performed their practice games as a team - although they alternated roles every other game so they controlled each half of the task four times during each practice session. All trainees' (both AIM-dyad and Individual-based) verbal communications and interactions were unobtrusively recorded. At the end of the experiment, trainees were informed of this and were given the option of allowing us to use their tapes or to erase the tapes. All trainees gave us permission to use the audio recordings for research purposes. This data is not presented or used in the presented study.

Due to administrative difficulties encountered in running some trainees, 22 trainees (8 from the AIM-dyad condition and 14 from the individual condition) from the final sample of 89 were unable to complete the eleventh experimental session which consisted of two standard test games of Space Fortress. However, there were no significant differences on any of the subsequent Space Fortress sessions between those trainees who did not complete Session 11 and those who did. These analyses were conducted as one-tail tests with the hypothesis that trainees who completed Session 11 would have higher scores because they had one extra session playing Space Fortress. There were also, no significant training protocol by completion-of-Session 11 (2x2) interactions on any of the subsequent Space Fortress sessions. The data for the 22 trainees who did not complete Session 11 were therefore considered to be similar to that for those who did. Further analyses also indicate that the two conditions did not differ in performance on Session 11 ($F[1, 69] = 1.28, p = 0.2628$). Consequently, in order to retain the same sample size across all sessions, Session 11 was excluded from all the analyses presented in this report.

RESULTS

Descriptive statistics, along with internal consistency estimates for the predictors and criteria used in the study, are presented in Table 3. Test-retest reliabilities were obtained for those predictors that were administered on multiple occasions. A test-retest reliability of 0.70 was obtained for the CA-VAT. For the three administrations of the Declarative Knowledge Test, the correlations between the first and second, and the first and third administrations were 0.61 and 0.45 (p = 0.001), respectively. The correlation between the second and third administrations was 0.75 (p = 0.001).

Table 3

Descriptive Statistics for Study Variables and Internal Consistency Estimates for Predictors - Total Sample

Measure	Mean	SD	Min	Max	Internal Consistency
PREDICTORS					
Aiming Task	2435.02	802.41	893.33	4973.33	0.84
CA-VAT	25.81	20.52	2.00	106.00	0.88
Declarative Knowledge					
First Admin.	27.47	2.19	19.00	30.00	0.63
Second Admin.	28.10	1.26	24.00	30.00	0.27
Third Admin.	27.33	1.59	22.00	30.00	0.33
Figure Matrices	859197.93	111453.07	590566.00	1093524.00	0.48
Processing Speed	306723.07	28446.04	214452.00	340902.00	0.76
Working Memory	420125.51	93493.26	239703.00	606742.00	0.71
CRITERIA					
SPACE FORTRESS					
Session 0	-1432.07	979.56	-3733.50	1619.50	---
Session 1	660.11	1485.58	-2398.50	3973.00	---
Session 2	1420.82	1552.15	-1871.50	4729.00	---
Session 3	1919.67	1642.77	-1423.50	4686.50	---
Session 4	2185.06	1676.92	-1124.50	5456.00	---
Session 5	2685.87	1675.35	-1420.50	5821.00	---
Session 6	2981.84	1658.67	-488.50	5982.00	---
Session 7	3215.13	1712.23	-476.00	6499.00	---
Session 8	3641.27	1620.82	-688.50	6333.00	---
Session 9	3807.09	1600.14	-59.00	6421.50	---
Session 10	3126.51	1527.67	-563.00	5775.00	---
[a]Session 12	701.53	1430.75	-3342.50	3498.00	---

Table 3 Continued

Measure	Mean	SD	Min	Max	Internal Consistency
Session 13	3674.40	1570.85	-547.00	6135.00	---
Session 14	3875.99	1568.66	-284.00	6534.50	---
Session 15	4066.91	1518.06	-495.00	6511.00	---
Session 16	4183.44	1583.97	-353.00	6973.50	---
Session 17	4195.06	1672.99	-510.00	6845.00	---
[A]Session 18	1596.45	1655.30	-3747.00	4265.00	---
Session 19	4340.87	1534.14	-747.00	6592.00	---
[B]Skill Loss	680.58	857.77	-1829.50	3172.00	---
STRATEGY					
Strategy 0	13.69	3.38	6.00	20.00	---
Strategy 1	15.66	3.85	6.00	22.00	---
Strategy 2	16.66	2.97	8.00	22.00	---
Strategy 3	16.27	3.52	4.00	22.00	---
Strategy 4	16.71	3.17	8.00	22.00	---
Strategy 5	17.61	3.25	10.00	22.00	---
Strategy 6	17.12	3.24	10.00	22.00	---
Strategy 7	18.53	2.80	9.00	22.00	---
Strategy 8	18.10	2.92	9.00	22.00	---
Strategy 9	18.41	2.53	12.00	22.00	---
Strategy 10	18.53	2.44	12.00	22.00	---
Strategy 13	19.23	2.41	8.00	22.00	---
Strategy 14	18.56	2.91	11.00	22.00	---
Strategy 15	19.34	2.94	10.00	22.00	---
Strategy 16	19.11	2.52	10.00	22.00	---
Strategy 17	18.65	2.68	9.00	22.00	---
Strategy 19	18.78	2.75	9.00	22.00	---
[B]Strategy	-0.12	2.08	-5.07	6.00	---

Table 3 Continued

Measure	Mean	SD	Min	Max	Internal Consistency
TRANSFER TASKS					
ASTEROIDS					
Asteroids 1	16890.28	7378.87	3860.00	49400.00	---
Asteroids 2	18822.35	6558.55	6480.00	48899.00	---
Asteroids 3	19397.00	4444.86	9140.00	32660.00	---
Asteroids 4	20017.67	3971.37	8200.00	28010.00	---
TEMPEST					
Tempest 1	27648.74	22449.98	3600.00	214692.00	---
Tempest 2	33457.67	17049.56	12449.00	149934.00	---
Tempest 3	38017.63	17007.64	15994.00	164359.00	---
Tempest 4	39757.32	15991.27	19089.00	150395.00	---

NOTE: [a]Keyboard Version of Space Fortress. *N* = 89. [a]Skill loss and strategy loss were operationalized as the performance difference between Session 9, the last acquisition trial and Session 10, the first session after the eight-week non-practice retention interval.

Table 4 presents descriptive statistics and *d* effect size differences between the AIM-dyad and Individual-based protocols on the predictors and criteria. Univariate significance tests for differences are also presented.

Table 4

Descriptive Statistics and Effect Size Differences Between Protocols on Predictors and Criteria

	AIM-DYAD		Individual-based		
Measure	Mean	SD	Mean	SD	d
PREDICTORS					
Aiming Task	2399.42	847.58	2464.08	771.23	-0.08
CA-VAT	26.33	24.11	25.39	17.31	0.05
Declarative Knowledge					
First Admin.	27.03	2.28	27.84	2.07	-0.37
Second Admin.	28.02	1.25	28.16	1.28	-0.11
Third Admin.	27.40	1.37	27.27	1.77	0.08
Figure Matrices	869748.63	108626.88	850585.12	114094.03	0.17
Processing Speed	309962.55	25070.30	304078.59	30934.37	0.21
Working Memory	427688.98	99346.82	413951.24	88990.77	0.15

Table 4 Continued

Measure	AIM-DYAD Mean	AIM-DYAD SD	Individual-based Mean	Individual-based SD	d
CRITERIA					
SPACE FORTRESS					
Session 0	-1468.12	1028.34	-1402.64	947.62	-0.07
Session 1	501.69	1609.12	789.44	1379.96	-0.19
Session 2	1192.53	1641.27	1607.19	1466.04	-0.27
Session 3	1684.41	1723.60	2111.72	1565.26	-0.26
Session 4	1997.26	1720.95	2338.37	1641.86	-0.20
Session 5	2473.43	1638.35	2859.29	1701.94	-0.23
Session 6	2738.50	1638.88	3180.48	1664.91	-0.27
Session 7	3004.83	1734.57	3386.82	1692.12	-0.22
Session 8	3509.30	1479.37	3749.00	1735.43	-0.15
Session 9	3648.27	1521.68	3936.74	1665.71	-0.18
Session 10	3028.24	1507.14	3206.73	1555.12	-0.12
[A]Session 12	929.25	1385.67	515.64	1454.07	0.29
Session 13	3671.96	1444.13	3676.40	1682.08	0.00
Session 14	3812.89	1633.80	3927.51	1528.54	-0.07
Session 15	4000.74	1285.77	4120.93	1695.60	-0.08
Session 16	4076.53	1528.19	4270.72	1638.59	-0.12
Session 17	4049.98	1673.78	4313.49	1680.23	-0.16
[A]Session 18	1725.87	1680.79	1490.80	1643.98	0.14
Session 19	4372.29	1372.10	4315.21	1668.45	0.04
[B]Skill Loss	620.03	876.53	730.01	848.00	-0.13
STRATEGY					
Strategy 0	14.59	2.88	12.95	3.60	0.49
Strategy 1	15.28	3.97	15.98	3.77	-0.18
Strategy 2	16.35	3.00	16.92	2.96	-0.19
Strategy 3	16.23	3.50	16.31	3.57	-0.02
Strategy 4	16.51	3.60	16.88	2.80	-0.12
Strategy 5	17.30	3.58	17.86	2.96	-0.17
Strategy 6	17.00	3.24	17.22	3.26	-0.07
Strategy 7	18.13	2.97	18.86	2.63	-0.26
Strategy 8	18.48	2.63	17.80	3.13	0.23
Strategy 9	18.57	2.23	18.29	2.76	0.11
Strategy 10	19.00	2.21	18.15	2.57	0.35
Strategy 13	19.78	1.73	18.79	2.78	0.41
Strategy 14	18.95	2.70	18.24	3.05	0.24
Strategy 15	19.86	2.42	18.92	3.27	0.32
Strategy 16	19.33	2.26	18.94	2.73	0.15
Strategy 17	18.83	2.10	18.51	3.09	0.12
Strategy 19	18.92	2.18	18.65	3.16	0.10
[B]Strategy	-0.43	2.26	0.14	1.90	-0.27

Table 4 Continued

Measure	AIM-DYAD Mean	AIM-DYAD SD	Individual-based Mean	Individual-based SD	d
TRANSFER TASKS					
ASTEROIDS					
Asteroids 1	15497.13	5521.93	18027.55	8491.67	-0.34
Asteroids 2	19636.40	6658.20	18157.82	6468.19	0.23
Asteroids 3	19251.75	4199.79	19515.57	4675.23	-0.06
Asteroids 4	19845.71	3914.00	20158.04	4052.54	-0.08
TEMPEST					
Tempest 1	24985.85	9253.92	29822.52	29046.37	-0.22
Tempest 2	30030.97	10091.79	36254.97	20794.35	-0.37
Tempest 3	35900.90	7812.26	39745.58	21770.16	-0.23
Tempest 4	39466.95	13879.55	39994.35	17668.39	-0.03

NOTE: [a]Keyboard Version of Space Fortress. The effect size statistic, *d*, is the standardized difference between two means. In computing *d*, AIM-dyad is the "experimental" condition ($n = 40$) and the Individual-based condition is the "control" ($n = 49$). *Univariate test for differences is significant at $p < .05$. All tests are two-tailed. [a]Skill loss and strategy loss were operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10).

Operationalization of Skill Acquisition, Retention and Loss, and Re-acquisition

The purpose of this section is to present a note identifying the Space Fortress sessions used to operationalize skill acquisition, retention and loss, and re-acquisition. Session 0 is the baseline. Sessions 1-9 are the acquisition sessions. Session 10 is the first session after the eight-week non-practice interval. This was a test-only session without any practice; it is considered to be a measure of retention. Consequently, skill loss was operationalized as the difference between Session 9 and Session 10. As previously noted, Session 11 was excluded from the analyses due to the unavailability of data for 22 (i.e., 25%) of the 89 trainees. Sessions 12 and 18 represent the keyboard version of Space Fortress, which was used as a transfer task. And finally, Sessions 13-17, and 19 represent re-acquisition sessions. Although Session 13 was also a test-only session, it was considered to be a re-acquisition session because Session 10 in effect served as a "practice" session for this session.

Tests of the research questions are presented below.

Will the amount and rate of skill acquisition be the same for the AIM-dyad and Individual-based training protocols?

The first research objective was to test for differences between the AIM-dyad and Individual-based protocols in terms of the amount of skill acquisition attained by trainees in the two protocols. Using a mixed factorial analysis of variance (ANOVA), results of a between subjects main effect ANOVA indicated that the training protocols did not result

in different levels of Space Fortress performance during the acquisition phase (F [1, 87] = 1.16, p = 0.2844). Furthermore, although results of a within subjects main effect ANOVA indicated a significant session effect (F[9, 783] = 452.63, p = 0.0001), the training protocols were not differentially effective in improving performance over sessions (i.e., the condition-by-session interaction term was not significant, F[9, 783] = 0.58, p = 0.8109). Thus both the amount and rate of skill acquisition appeared to be the same for the AIM-dyad and Individual-based protocols. The above results are further illustrated in Figure 2 which presents the performance of the dyads and individuals across Space Fortress sessions.

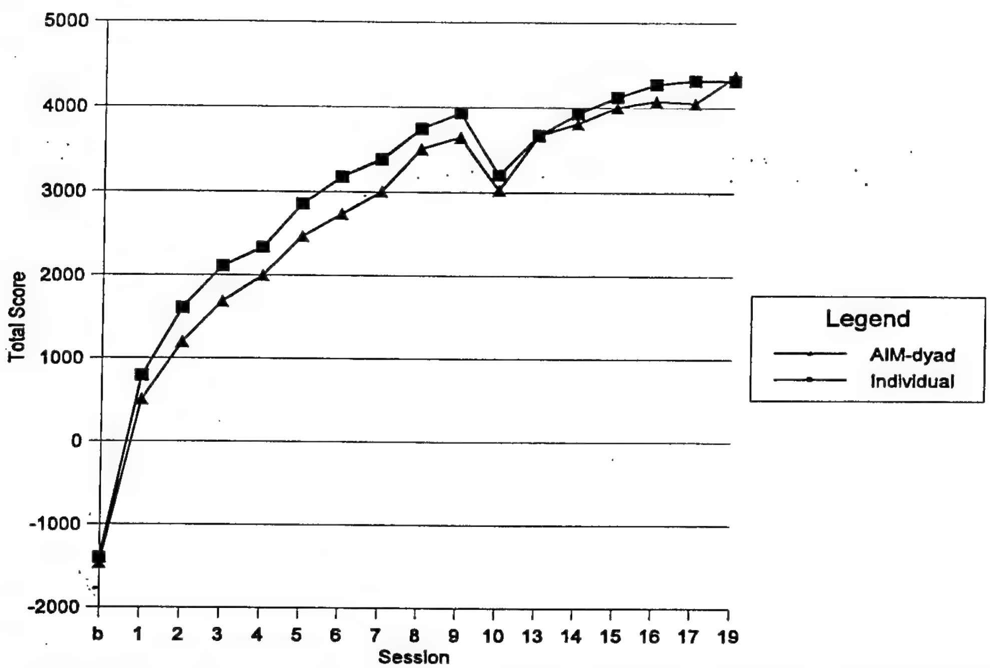

Figure 2. Mean total score on two test games of Space Fortress as a function of session for AIM-dyad and individual conditions.

Will the AIM-dyad and Individual-based protocols result in different amounts of skill loss?

Any discussion of skill acquisition and subsequent loss must be preceded with an assurance that the skill in question was mastered or, at the very least, that a reasonable amount of acquisition had taken place. In the absence of this, any attempts to demonstrate re-acquisition may simply represent an increased amount of skill acquisition. Because the present study was not able to train trainees to a specified level of mastery due to the use of a dyadic protocol which yoked the training of pairs of trainees, the assessment of task mastery level was accomplished on a post hoc basis.

Specifically, to determine the extent to which trainees had *really* mastered the task, linear and curvilinear regression lines were fit to the mean acquisition data presented in Figure 2. A better fit for the curvilinear line would be indicative

of a decrease in the rate of performance improvement over sessions, suggesting that trainees may have been approaching mastery. On the other hand, a better fit for a linear line would suggest that trainees were clearly still in an acquisition phase. Results of these analyses demonstrated a better fit for the curvilinear line (F [2, 7] = 72.47, p = 0.0088, R^2 = 0.95) compared to the linear line (F[1, 8] = 52.98, p = 0.0001, R^2 = 0.87; R^2 increment = 0.08, p = 0.05). Thus although these data do not provide any information on the level of mastery obtained by any one single participant, as a group, the data indicates that the trainees may have approached mastery.

Skill loss was operationalized as the performance difference between Session 9, the last acquisition trial, and Session 10, the first session after the eight-week non-practice retention interval. As shown in Table 4, although both conditions experienced some skill loss (d = 0.41 [t = 4.47, p = 0.0001] and 0.45 [t = 6.03, p = 0.0001] for dyads and individuals, respectively), results of a one-way ANOVA failed to obtain any significant differences *between* the two protocols in terms of the amount of skill loss (t = 0.06, p = 0.5504). However, in terms of d effect sizes, the Individual-based protocol displayed slightly more skill loss than the AIM-dyad protocol (d = -0.13).

Will the amount and rate of skill re-acquisition be the same for the AIM-dyad and Individual-based training protocols?

The analyses performed to answer this question were similar to those for the acquisition phase. Again, results of a between subjects main effect ANOVA indicated that the training protocols did not result in different levels of Space Fortress performance during the re-acquisition phase (F[1, 87] = 0.11, p = 0.7389). Results of a within subjects main effect ANOVA indicated a significant session effect (F [5, 435] = 16.21, p = 0.0001). Finally, like the acquisition phase, the training protocols were not differentially effective in improving performance over sessions (i.e., the condition-by-session interaction term was not significant, F[5, 435] = 0.97, p = 0.4380). Thus both the amount and rate of skill re-acquisition appeared to be the same for the AIM-dyad and Individual-based protocols. These results are further illustrated in Figure 2.

Supplementary Analysis - Practice Sessions

Additional analyses were performed to compare dyadic to individual performance on practice trials. Table 5 presents descriptive statistics and d effect size differences between the AIM-dyad and Individual-based protocols on the practice sessions. Univariate significance tests for differences are also presented. Figure 3 shows the average total scores on the eight practice games for both groups over both acquisition and re-acquisition sessions. This chart clearly indicates that dyads performed increasingly better than individuals during practice sessions. Although the overall analysis of conditions by practice over sessions failed to obtain a significant main effect for conditions for both acquisition (F [1, 87] = 1.63, p = 0.2057) and re-acquisition (F [1, 87] = 3.63, p = 0.0602), the main effects for sessions for acquisition (F [8, 696] = 381.38, p = 0.0001) and re-acquisition (F [3, 261] = 9.48, p = 0.0001) were

significant. The condition by session effects were also significant ($F[8, 696] = 2.73, p = 0.0057$ for acquisition; and $F[3, 261] = 2.86, p = 0.0373$ for re-acquisition).

These results indicate that, during acquisition, dyads performed increasingly better than individuals during practice; a finding also reported by Shebilske et al. (1992). This finding is consistent with the cognitive complexity and information processing demands required by Space Fortress and, in fact, it comes as no surprise that "two heads are better than one" in the performance of this task.

Table 5

Descriptive Statistics and Effect Size Differences Between Protocols on Space Fortress Practice Sessions

	AIM-DYAD		Individual-based		
Space Fortress Practice Session	Mean	SD	Mean	SD	*d*
Session 1	229.82	1132.12	180.44	1123.06	0.04
Session 2	1428.58	1169.41	1202.11	1398.14	0.17
Session 3	1904.72	1272.47	1783.75	1495.80	0.09
Session 4	2364.47	1189.58	2057.84	1595.08	0.21
Session 5	2777.17	1236.84	2370.11	1577.90	0.28
Session 6	3248.05	1302.19	2747.91	1560.77	0.34
Session 7	3394.04	1275.14	3003.83	1607.31	0.27
Session 8	3873.81	1121.38	3213.43	1670.65	0.45*
Session 9	3989.98	1034.48	3509.30	1605.01	0.35
Session 14	4289.15	1078.47	3582.17	1601.45	0.50*
Session 15	4374.67	958.13	3789.06	1697.28	0.40
Session 16	4499.34	943.22	3954.96	1611.92	0.40
Session 17	4387.31	1086.53	4040.55	1626.97	0.25

NOTE: The effect size statistic, *d*, is the standardized difference between two means. In computing *d*, AIM-dyad is the "experimental" condition ($n = 40$) and the Individual-based condition is the "control" ($n = 49$). *Univariate test for differences is significant at $p < .05$. All tests are two-tailed.

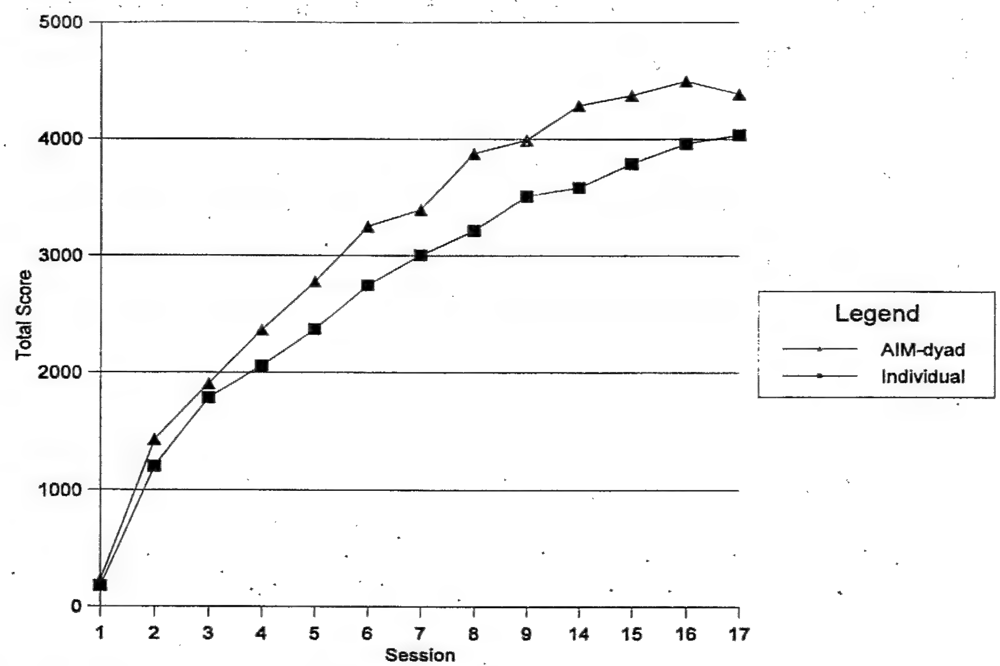

Figure 3. Total score on Space Fortress averaged over eight practice games as a function of session for AIM-dyad and individual conditions.

Relationship Between Ship Control Strategy and Space Fortress Performance

A second dependent variable used in the present study was ship control strategy. This variable was operationalized as the proctor's rating of the trainees performance on four control strategy variables, namely ship maneuvering strategy, ship speed, joystick manipulation, and course deviation due to mines. We first sought to assess the psychometric adequacy of the strategy ratings. This was accomplished by looking at the inter-rater reliability and accuracy of the ratings.

Inter-rater reliability was computed for a subset of raters. Specifically these were the seven proctors who had participated in the first of the three phases of data collection. The reliability data was collected using procedures described in the ***Proctor Strategy Rating Training*** section of this report (i.e., these raters rated the same Space Fortress sessions during the proctor training session). Using coefficient alpha as a measure of intraclass correlations (Hays, 1988; Winer, Brown, & Michels, 1991), the inter-rater reliability across the seven raters for the overall strategy score (i.e., the sum of the four components) was 0.92. The levels of inter-rater reliability for the four component scores were also very high. The results of the inter-rater reliability analyses are presented in Table 6.

Table 6

Results of Inter-Rater Reliability Analyses for Overall Strategy and Component Scores

Strategy Score	α
Overall Strategy	0.92
Ship-Maneuvering Strategy	0.90
Ship Speed	0.93
Joystick Manipulation	0.88
Course Deviation due to Mines	0.81

The accuracy of the strategy ratings was assessed by investigating hypothesized convergent and divergent relationships between specified strategy variables (ratings) and Space Fortress sub-scores. Two sets of similar scores which were expected to display a convergent relationship were ship speed strategy and velocity, and ship maneuvering strategy and control. Conversely, to assess the divergent validity of the strategy ratings, it was hypothesized that ship speed strategy and control, and ship maneuvering strategy and velocity would display divergent relationships. The correlations representing these relationships are presented in Table 7. The results clearly indicate that the hypothesized convergent and divergent predictions were supported. The mean correlations for ship speed strategy and velocity, and ship maneuvering strategy and control were 0.73 (*SD* = 0.03) and 0.62 (*SD* = 0.09) respectively. On the other hand, the mean correlations for ship speed strategy and control, and ship maneuvering strategy and velocity were 0.39 (*SD* = 0.07) and 0.39 (*SD* = 0.08) respectively. In summary these results, coupled with the inter-rater reliability data presented in Table 6, demonstrate that the proctors' ship control strategy ratings were reasonably accurate and reliable.

Table 7

Correlations Between Hypothesized Convergent and Divergent Strategy Ratings and Space Fortress Performance Sub-Scores

	Hypothesized Convergence		Hypothesized Divergence	
Session	SS/VEL	SM/CNTL	SS/CNTL	SM/VEL
Session 0	0.48	0.38	0.28	0.18
Session 1	0.79	0.59	0.38	0.43
Session 2	0.76	0.57	0.29	0.17
Session 3	0.68	0.43	0.23	0.35
Session 4	0.77	0.51	0.42	0.34
Session 5	0.72	0.51	0.34	0.36
Session 6	0.70	0.64	0.30	0.38
Session 7	0.76	0.47	0.41	0.50
Session 8	0.72	0.66	0.43	0.49
Session 9	0.72	0.64	0.47	0.39
Session 10	0.69	0.71	0.30	0.32
Session 13	0.69	0.69	0.45	0.36
Session 14	0.71	0.67	0.39	0.41
Session 15	0.77	0.70	0.47	0.40
Session 16	0.70	0.71	0.46	0.48
Session 17	0.75	0.69	0.48	0.46
Session 19	0.76	0.72	0.42	0.42
MEAN	0.73	0.62	0.39	0.39
SD	0.03	0.09	0.07	0.08

NOTE: SS = ship speed (strategy); SM = ship-maneuvering (strategy); VEL = velocity (Space Fortress performance); and CNTL = control (Space Fortress performance). Correlations greater than 0.18 are significant at $p < .05$; those greater than 0.25 are significant at $p < .01$; and those greater than 0.32 are significant at $p < .001$. All tests are one-tailed.

The next question of interest was to assess the relationship between ship control strategy and Space Fortress performance. It was expected that trainees who demonstrated more effective control strategies would have higher Space Fortress performance scores than those who did not. The correlations presented in Table 8 indicate that this hypothesis was supported; the relationships between strategy and Space Fortress performance were consistently moderate to high, with a mean effect of 0.60 (*SD* = 0.07) across both acquisition and re-acquisition sessions for corresponding strategy and space fortress session scores (i.e., the diagonal numbers in Table 8).

Table 8

Correlation Between Space Fortress Performance and Ship Control Strategy

SPACE FORTRESS SESSIONS	STRATEGY															
	0	1	2	3	4	5	6	7	8	9	10	13	14	15	16	17	19
0	**36**	41	53	41	30	45	42	35	30	29	36	21	24	20	29	18	27
1	24	**65**	56	54	53	62	54	45	38	41	44	36	42	31	36	36	45
2	23	58	**59**	53	60	58	55	49	41	49	42	34	43	29	33	36	43
3	15	53	59	**58**	60	64	61	55	44	46	41	34	46	31	35	37	41
4	18	51	57	56	**59**	63	60	56	44	44	44	36	45	32	40	35	39
5	15	54	58	52	57	**70**	62	57	50	49	46	37	49	40	36	41	42
6	05	50	59	52	51	67	**67**	65	53	50	49	38	54	41	46	49	49
7	12	49	53	40	49	66	60	**64**	55	47	51	40	51	43	43	47	44
8	11	49	59	42	44	63	58	65	**60**	53	55	45	59	50	48	55	53
9	09	47	55	49	46	65	55	69	59	**60**	58	49	54	49	50	58	55
10	12	45	54	38	41	61	56	72	57	56	**64**	51	51	51	50	56	53
13	18	47	62	38	47	63	54	65	61	61	65	**56**	60	55	57	57	57
14	07	46	59	42	47	62	56	61	62	46	60	51	**61**	51	51	63	60
15	04	40	54	46	43	60	61	64	59	51	59	50	62	**56**	54	62	60
16	04	41	53	47	42	60	56	68	63	51	60	53	63	57	**54**	64	65
17	07	45	52	50	36	56	56	68	60	49	60	47	56	52	49	**62**	60
19	09	43	49	46	45	59	56	66	65	54	62	50	66	55	54	64	**63**

NOTE: Decimals have been omitted. Correlations between Space Fortress performance and ship control strategy score for corresponding sessions are in **boldface** (i.e., diagonal). Correlations greater than 0.18 are significant at $p < .05$; those greater than 0.25 are significant at $p < .01$; and those greater than 0.32 are significant at $p < .001$. All tests are one-tailed.

Differences between AIM-dyad and Individual-based training protocols in ship control strategy in skill acquisition and re-acquisition phases

Consistent with the results obtained for performance on Space Fortress, the results of the between subjects main effect ANOVAs indicated no significant differences between the two training protocols on ship control strategy during the acquisition (F[1, 87 = 0.02, p = 0.9898) and re-acquisition phases (F[1, 87] = 1.68, p = 0.19890). However, results of the within subjects main effect ANOVAs indicated significant session effects (F[9, 783] = 29.56, p = 0.0001; F[5, 435] = 3.27, p = 0.0121, for acquisition and re-acquisition, respectively). And finally, the condition-by-session interaction was significant for the acquisition phase (F[9, 783] = 1.99, p = 0.0380) but not for the re-acquisition phase (F[5, 435] = 0.75, p = 0.5895. As shown in Table 4, because of significant differences between

individuals and dyads on baseline strategy ratings, the preceding analyses were repeated with an analysis of covariance (ANCOVA) covarying out the baseline strategy score. The results of these analyses indicated that the baseline rating was not a significant covariate. Consequently, the results obtained for the ANCOVA were similar to those obtained for the ANOVA. In summary, as Figure 4 illustrates, although dyads had less effective strategies during the early sessions of skill acquisition, they surpassed individuals in the latter sessions.

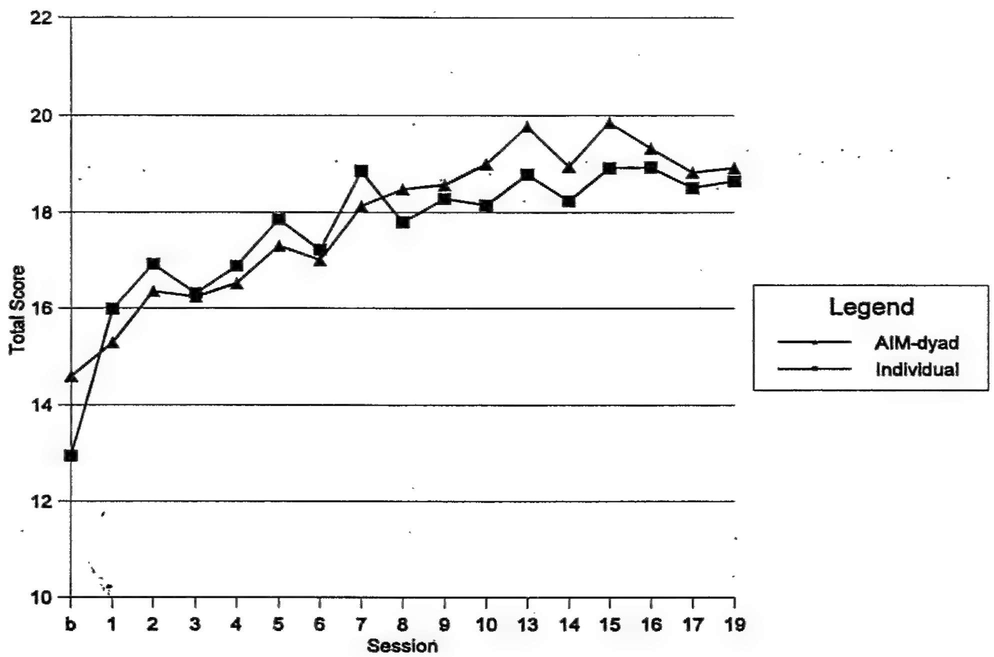

Figure 4. Mean strategy score as a function of session for AIM-dyad and individual conditions.

Effect of non-practice interval on ship control strategy for the AIM-dyad and Individual-based protocols

Strategy loss was operationalized as the difference in strategy scores for Session 9, the last acquisition trial, and Session 10, the first session after the eight-week non-practice retention interval. As shown in Table 4, although dyads displayed a small gain in strategy scores over the non-practice interval (d = -0.19), this effect was not significant (t = -1.21, p = 0.2331). On the other hand, there was no change in strategy scores for individuals from the acquisition to the retention session (d = 0.05, t = 0.52, p = 0.6053). Further analyses also failed to obtain any significant differences *between* the two protocols in terms of the amount of strategy loss (t = 1.30, p = 0.1962). However, in terms of d effect sizes, the Individual-based protocol displayed moderately more strategy loss than the AIM-dyad protocol (d = -0.27).

Figure 5 presents the mean score for each ship control strategy component over sessions. This figure demonstrates that although there were no consistent or patterned shifts in the relative magnitude of strategy sub-scores across sessions, there was a general improvement in overall strategy scores. However, more interesting results are apparent when comparing the pattern of strategy sub-component scores of the AIM-dyad protocol (Figure 6) to that of the Individual-based protocol (Figure 7). Although the ship-maneuvering strategy appears to be the predominant strategy in both protocols, one primary difference is evident within the remaining three sub-component strategies. For instance, the joystick manipulation strategy is typically the lowest of the sub-component scores in the AIM-dyad protocol, whereas ship speed strategy is typically the lowest sub-component score in the Individual-based strategy. These results lend some support to the argument that, due to a substantial reduction in hands-on practice, trainees in the AIM-dyad protocol would rely less on joystick control strategies and more on complex cognitive strategies compared to trainees in the Individual-based protocol.

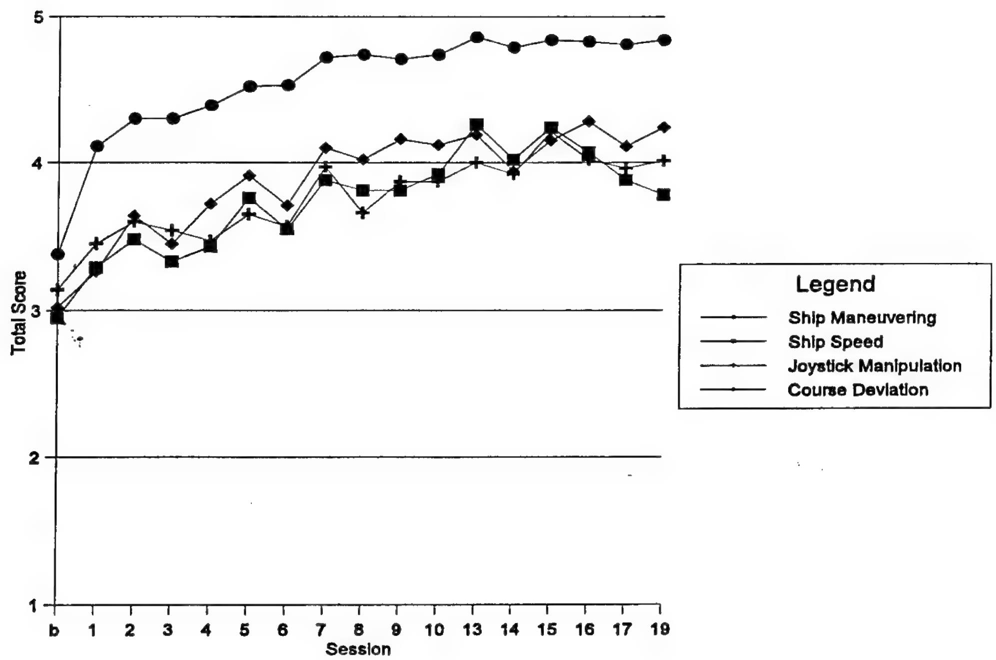

Figure 5. Strategy sub-component scores as a function of sessions.

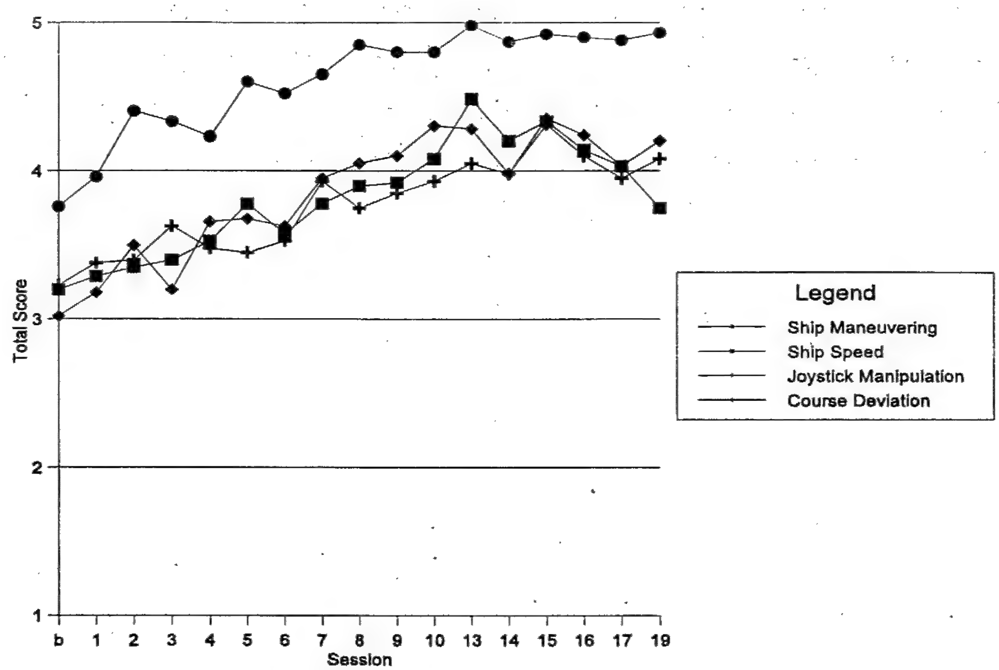

Figure 6. Strategy sub-component scores as a function of sessions for AIM-dyad condition.

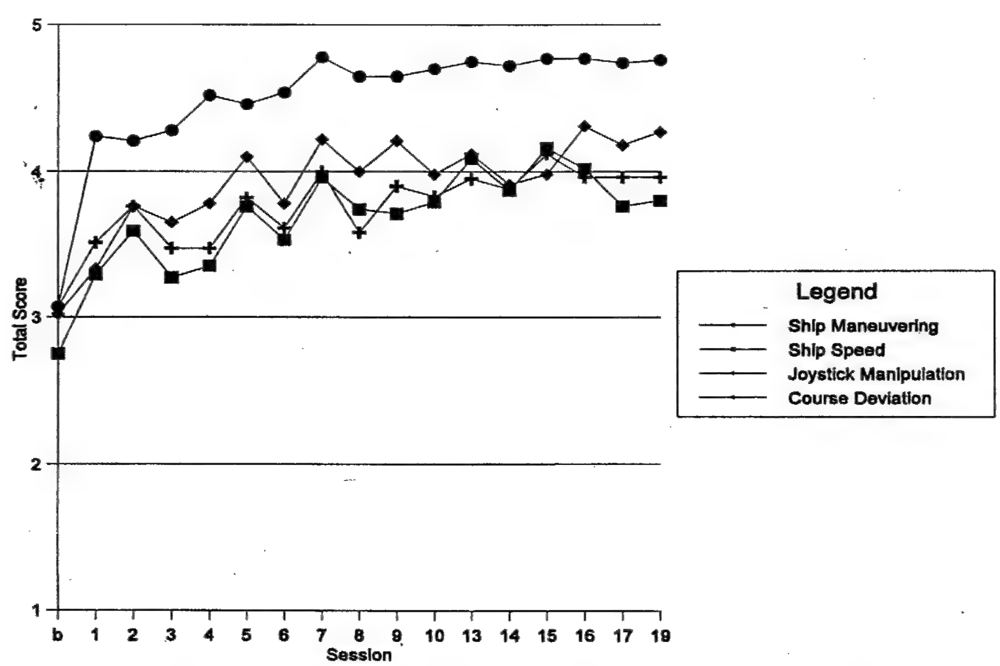

Figure 7. Strategy sub-component scores as a function of sessions for individual condition.

Differences between AIM-dyad and Individual-based training protocols on the transfer tasks

As previously noted, data were collected on trainees' self-reported ability levels and the extent to which they had played the two transfer tasks (i.e., Asteroids and Tempest) before the commencement of training and also after the 8-week non-practice interval. Analyses of these data indicated that trainees in the two protocols did not differ on any of these variables.

In the absence of any previous empirical data on the extent to which Asteroids and Tempest were satisfactory transfer tasks, the present study first sought to assess the relationship between these two tasks and Space Fortress. This was accomplished by computing the correlation between pre-training performance scores on these tasks and Space Fortress with the intention of using positive or negative relationships to infer the extent to which these tasks can be considered to be positive or negative transfer tasks.

On the basis of its graphic interface, procedural, and operational rules, it was expected that Asteroids would serve as a positive transfer task. Although the exact status of Tempest appeared to be more ambiguous, it was expected to serve and function as a negative transfer task primarily because its graphic interface, procedural and operational rules are very different from those of Space Fortress.

The correlations between the first (i.e., baseline) administration of Asteroids and Tempest and Space Fortress performance are presented in Table 9. These correlations indicate that, as hypothesized, although the magnitude of the effects were small, performance on Asteroids was positively related to Space Fortress performance. Specifically, trainees who scored better on Space Fortress also performed better on Asteroids. There appeared to be no relationship between Tempest and Space Fortress performance. So the use of Asteroids as a positive transfer task does not appear to be inappropriate; and at worst, Tempest can be considered to be a neutral task. The correlation between the baseline administrations of Asteroids and Tempest was moderate (r = 0.60, p = 0.0001).

Table 9

Correlation Between Baseline Administrations of Asteroids and Tempest and Space Fortress Performance

Space Fortress	Asteroids	Tempest
0	0.19	0.10
1	0.24*	0.16
2	0.21*	0.19
3	0.21*	0.16
4	0.22*	0.11
5	0.29**	0.15
6	0.27**	0.19
7	0.22*	0.16
8	0.28**	0.17
9	0.24*	0.17
10	0.17	0.04
[a]12	-0.10	-0.16
13	0.21*	0.10
14	0.18	0.11
15	0.25*	0.13
16	0.19	0.09
17	0.19	0.10
[a]18	-0.02	-0.24*
19	0.20	0.11

NOTE: [a]Keyboard version of Space Fortress.
**p* < .05, ***p* < .01. All tests are two-tailed.

Although the transfer research objective was primarily exploratory in nature, it was expected that trainees who learned the complex cognitive strategies underlying high performance of the Space Fortress task would successfully transfer these strategies to the keyboard version and Asteroids from the normal version. Consequently, because AIM-dyad trainees may develop more complex cognitive strategies due to their reduced hands-on practice, it was expected that they would perform better than Individual-based trainees on the transfer tasks in the acquisition, retention, and re-acquisition phases of training.

Descriptive statistics along with *d* effect size differences between the AIM-dyad and Individual-based protocols on the three transfer tasks are presented in Table 4. Univariate significance tests for differences are also presented. For Asteroids, results of a between subjects main effect ANOVA failed to reveal a significant difference between the AIM-dyad and Individual-based protocols (F [1, 87] = 0.25, p = 0.6193). Results of a within subjects main effect ANOVA indicated a significant session effect (F [3, 261] = 7.11, p = 0.0001). However, the session-by-condition interaction was not significant (F [3, 261] = 2.42, p = 0.0664). In terms of effect sizes, the results indicate that dyads performed better than individuals on the second administration of Asteroids (d = 0.23). However, the performance differences on the other two post-baseline administrations of Asteroids (Asteroids 3 and 4) were negligible (d = -0.06 and -0.08). Figure 8 presents performance means on the four administrations of Asteroids by condition. As shown in both this figure and Table 4, there were relatively large differences between individuals and dyads on the first (baseline) administration of Asteroids. Consequently, the preceding analyses were repeated with an ANCOVA covarying out the first administration. And, although the covariate was significant (F [1, 86] = 8.70, p = 0.0041), the results obtained for the ANCOVA were similar to those obtained for the ANOVA. Specifically, neither condition (F [1, 86] = 0.79, p = 0.3776), session (F [2, 172] = 0.73, p = 0.4816), or session-by-condition effects (F [2, 172] = 1.34, p = 0.2657) were significant.

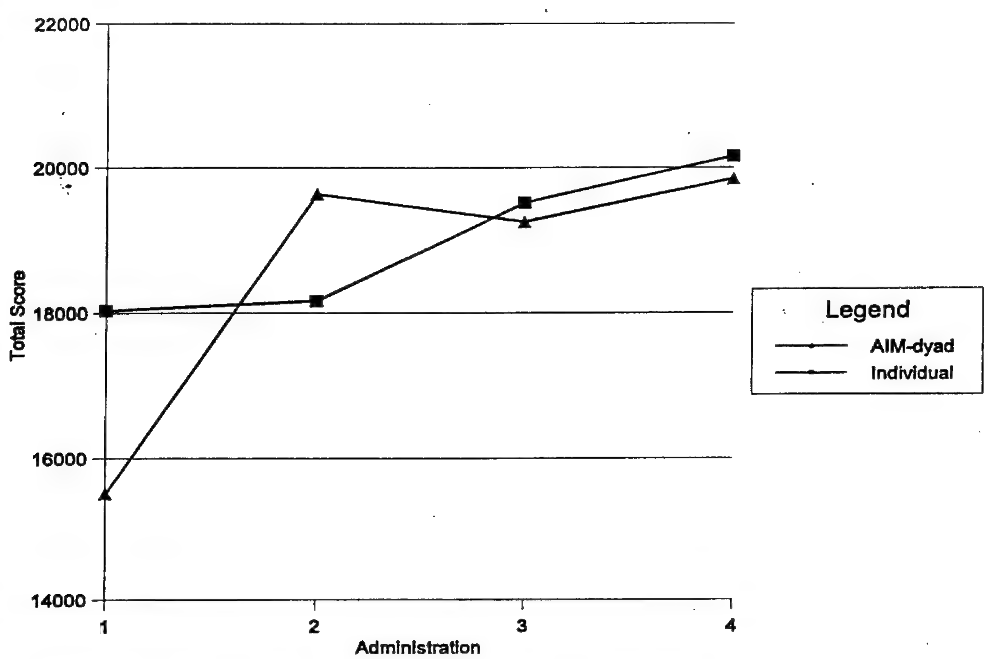

Figure 8. Asteroids score as a function of administration for AIM-dyad and individual conditions.

The results obtained for Tempest were similar to those for Asteroids. The main effect for condition was not significant (F [1, 87] = 2.66, p = 0.1068). And although the session effect was significant (F [3, 261] = 9.48, p =

0.0001), the session-by-condition interaction was not ($F[3, 261] = 0.46, p = 0.7093$). The effect sizes indicated that on the second and third administrations of Tempest, individuals performed better than dyads (d = -0.37 and -0.23, respectively). The performance difference on the third post-baseline administration of Tempest (Tempest 4) was negligible (d = -0.03). Figure 9 presents performance means on the four administrations of Tempest by condition. Again, as with Asteroids, there were relatively large differences between individuals and dyads on the first (baseline) administration of Tempest. Consequently, the preceding analyses were repeated with an ANCOVA covarying out the first administration. However, although the covariate was significant ($F[1, 86] = 5.70, p = 0.0241$), the results obtained for the ANCOVA were similar to those obtained for the ANOVA. Specifically, neither condition ($F[1, 86] = 1.62, p = 0.2070$), session ($F[2, 172] = 2.86, p = 0.0599$), or session-by-condition effects ($F[2, 172] = 0.69, p = 0.5024$) were significant.

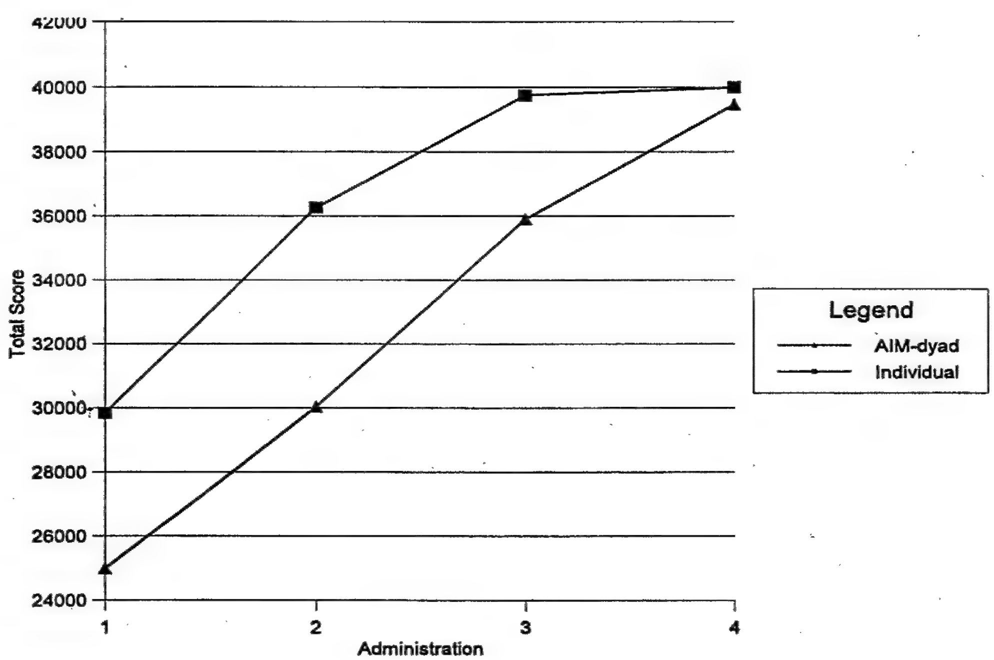

Figure 9. Tempest score as a function of administration for AIM-dyad and individual conditions.

Finally, for the keyboard version of Space Fortress, the main effect for condition was not significant ($F[1, 87] = 1.12, p = 0.2922$). And again, although the session effect was significant ($F[1, 87] = 52.49, p = 0.0001$), the session-by-condition interaction was not ($F[1, 87] = 0.53, p = 0.4673$). Like the other positive transfer task, Asteroids, the effect sizes indicate that dyads performed better than individuals on the two administrations of the keyboard version of Space Fortress (d = 0.29 and 0.14, respectively). Figure 10 presents performance means on the two administrations of the keyboard version of Space Fortress, along with the standard version, by condition.

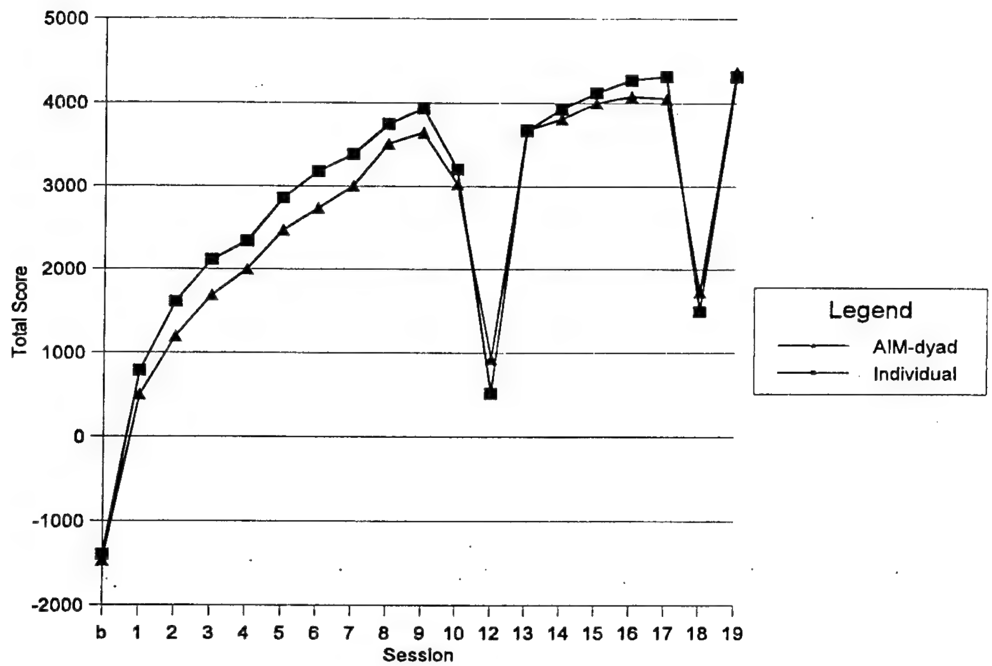

Figure 10. Mean total score on two test games of Space Fortress and Keyboard sessions as a function of session for AIM-dyad and individual conditions.

Figure 10 shows that there appears to be a slight crisscross in the performance on the original and keyboard versions of Space Fortress scores for the two conditions. Specifically, the AIM-dyad protocol scores appear to be slightly lower than the Individual-based scores across normal versions of Space Fortress (mean d = -.14). However, on the keyboard version, the AIM-dyad scores appear to be slightly higher than Individual-based scores (mean d = .22). In summary, although the results of the significance tests were not supportive of the study hypotheses, the effect sizes of the dyad and individual performance differences on the transfer tasks were more consonant with the hypotheses. In general, dyads appeared to perform better on the positive transfer tasks (i.e., Asteroids and Space Fortress Keyboard version), and worse on the neutral transfer task (i.e., Tempest).

What is the nature of the ability (i.e., cognitive ability, declarative knowledge, psychomotor ability, spatial processing speed, spatial working memory, and visual attention) and performance relationships over the acquisition, loss, and re-acquisition phases of task performance? Will these relationships be influenced by the training condition to which trainees were assigned?

The goal of this research objective was to assess the magnitude of the relationship between specified ability variables and Space Fortress performance, and also the stability of these relationships (i.e., increase, decline, or flat) over the acquisition, loss, and re-acquisition phases of task performance. The effect of training protocol was also assessed. The present study investigated the effects of six ability variables. The results of each ability variable are presented

in turn. Correlations matrices representing the relationships amongst predictors and criteria can be found in Appendices A, B, and C.

Cognitive Ability (Figure Matrices Test). There were several objectives to the analyses presented here. The first was to determine whether individual differences in cognitive ability would predict Space Fortress performance. The second was to assess the stability of the cognitive ability-performance relationship. The results presented in Table 10 indicate that *g* was moderately related to Space Fortress performance across all sessions. For the total sample, the mean correlation for the acquisition phase was 0.30 (SD = 0.06). The mean correlations for dyads and individuals were 0.29 (SD = 0.07) and 0.30 (SD = 0.06) respectively. Similar effects were obtained for the re-acquisition phase (r = 0.21 [SD = 0.05]; r = 0.21 [SD = 0.05]; r = 0.21 [SD = 0.06] for the total sample, dyads, and individuals, respectively).

Table 10

Correlations Between Cognitive Ability and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	A I M - Dyad	Individual-based
Session 0	0.20	0.13	0.25
Session 1	0.35***	0.32*	0.40**
Session 2	0.28**	0.34*	0.26
Session 3	0.33**	0.36*	0.33*
Session 4	0.35***	0.37*	0.37**
Session 5	0.31**	0.33*	0.32*
Session 6	0.27*	0.35*	0.24
Session 7	0.26*	0.25	0.29*
Session 8	0.24*	0.27	0.24
Session 9	0.29*	0.22	0.30
Session 10	0.25*	0.21	0.30*
Session 13	0.24*	0.20	0.27
Session 14	0.18	0.19	0.18
Session 15	0.19	0.29	0.14
Session 16	0.20	0.25	0.18
Session 17	0.25*	0.20	0.30*
Session 19	0.18	0.15	0.21
[a]Skill Loss	0.03	0.04	0.04

Table 10 Continued

Space Fortress Session	Total Sample	A I M - Dyad	Individual-based
TRANSFER TASKS			
[A]Session 12	0.34**	0.34*	0.33*
[A]session 18	0.24*	0.26	0.22
Asteroids 1	-0.06	-0.03	-0.05
Asteroids 2	0.01	-0.10	0.07
Asteroids 3	0.10	0.05	0.15
Asteroids 4	0.13	-0.05	0.27
Tempest 1	-0.09	0.07	-0.13
Tempest 2	0.10	0.03	0.16
Tempest 3	0.13	0.24	0.13
Tempest 4	0.11	0.09	0.14

NOTE: [A]Keyboard Version of Space Fortress. [a]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). $*p < .05$; $**p < .01$; $***p < .001$. All tests are two-tailed.

Figure 11 presents the overall *g*-performance correlations and the *g*-performance correlations for each training protocol across all training sessions. To assess the stability of the *g*-performance relationship hierarchical regression analysis was performed where the *g*-performance correlation was regressed on session, session squared, training condition, condition-by-session interaction, and condition-by-session squared interaction. A significant session effect indicates a linear relationship between the *g*-performance correlation and amount of training. A positive linear trend would thus indicate that *g* becomes more strongly related to performance in later training sessions. A significant session squared effect would indicate a curvilinear trend whereby the linear trend between session and the *g*-performance correlation is asymptotic. A significant effect for training condition would indicate that averaged over all sessions, the *g*-performance relationship was stronger for one training condition, compared to the other. It is important to note that training condition was a dichotomous variable where the dyad condition was coded a 1 and the individual condition as 0. A significant condition-by-session interaction effect would indicate that the relationship between the *g*-performance correlation and session is linear for one training protocol but not for the other. Lastly, a significant condition-by-session squared interaction would indicate that a curvilinear trend between session and the *g*-performance correlation exists for one training protocol but not for the other.

The results of hierarchical regressions to test the fit of linear and curvilinear lines to the data which are presented in Table 11, indicate that there was a linear decrease in the magnitude of the relationship between *g* and performance over sessions. No other incremental effects were significant.

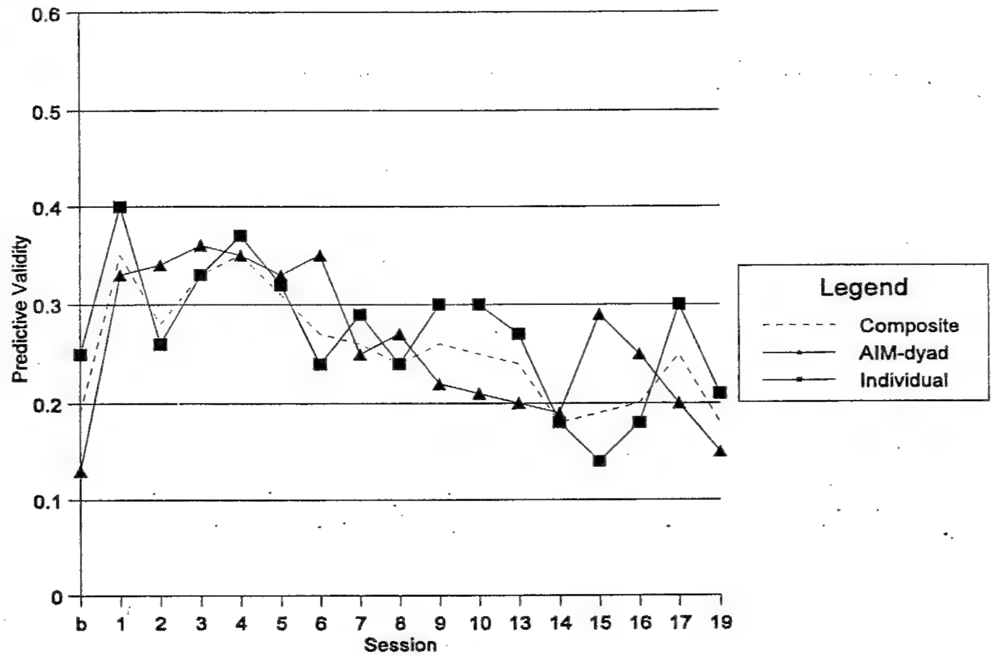

Figure 11. Cognitive ability-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 11

Hierarchical Regression of g*-Performance Correlations on Session and Condition to Test for Nature of fit*

Models	[a]B	Model R^2	ΔR^2
Session	-0.0039	0.2827**	
Session^2	-0.0002	0.3267**	0.0440
Training Condition	-0.0693	0.3319**	0.0052
Condition * Session	0.01	0.3327*	0.0008
(Condition * Session)^2	-0.0009	0.3525*	0.0198

NOTE: AIM-dyad = 1; Individual-based = 0. [a]The numbers presented are from the final regression equation. *$p < .05$; **$p < .01$; ***$p < .001$.

To further examine the nature of the *g*-performance relationship, differences between high and low *g* trainees on Space Fortress sessions were assessed. To accomplish this, a median split on the Figure Matrices test was used to create high and low *g* groups. Results of a 2x17 (*g* x practice) mixed factors ANOVA indicated that practice did not eliminate performance differences between high and low *g* trainees (see Figure 12). The results of these analyses indicated that there was a significant effect for *g* (F[1, 87] = 7.73, p = 0.0067), and a significant practice effect (F[16, 1392] =

398.35, p = 0.0001). However, the g-by-practice interaction was not significant (F[16, 1362] = 0.98, p = 0.4486). Therefore, the results indicate that the rate of improvement on Space Fortress was the same for both high and low g trainees.

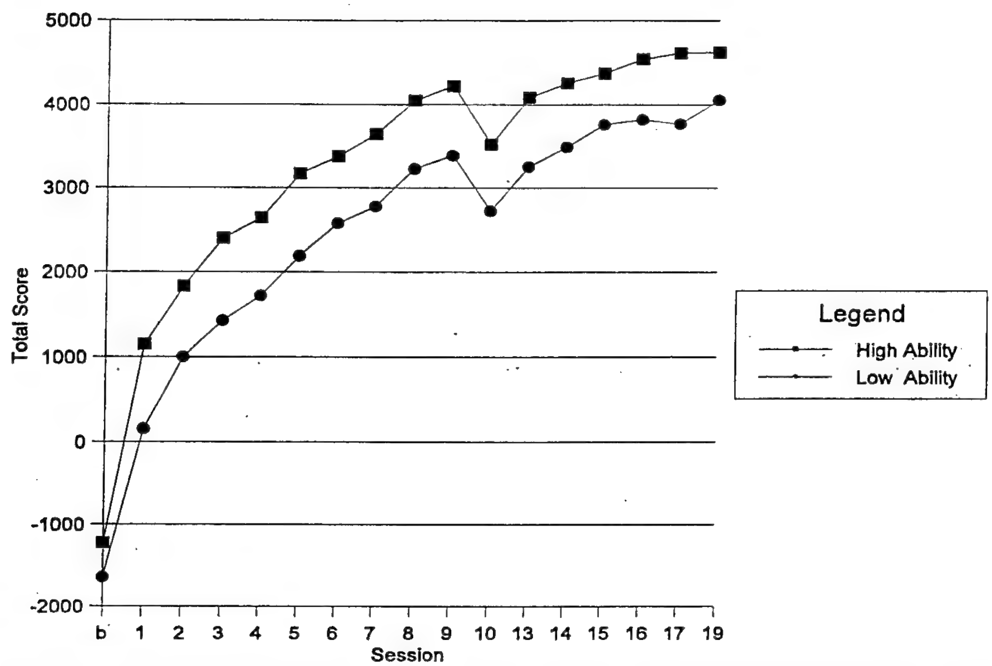

Figure 12. High and low cognitive ability trainees' Space Fortress performance across training sessions.

Declarative Knowledge. The objectives and analyses for this variable were identical to those previously presented for cognitive ability. Although the Declarative Knowledge Test was administered on three occasions, the analyses presented here are based on only the first administration. The primary reason for this is that the first administration was the most comparable to the other individual difference measures in terms of when it was administered in reference to the training protocols. That is, like the other measures, it was administered before the trainees commenced training.

Table 12

Correlations Between Declarative Knowledge (First Administration) and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	AIM-Dyad	Individual-based
Session 0	0.33**	0.22	0.43**
Session 1	0.40***	0.40*	0.38**
Session 2	0.35***	0.38*	0.29*
Session 3	0.40***	0.46**	0.31*
Session 4	0.36***	0.47**	0.24
Session 5	0.35***	0.36*	0.32*
Session 6	0.30**	0.31	0.26
Session 7	0.27*	0.24	0.27
Session 8	0.32**	0.28	0.34*
Session 9	0.32**	0.39*	0.25
Session 10	0.37***	0.37*	0.36*
Session 13	0.36***	0.40**	0.34*
Session 14	0.28**	0.28	0.27
Session 15	0.34**	0.41**	0.30*
Session 16	0.35***	0.36*	0.34*
Session 17	0.31**	0.32*	0.28
Session 19	0.29**	0.31	0.30*
[B]Skill Loss	-0.06	0.04	-0.18
TRANSFER TASKS			
[A]Session 12	0.31**	0.37*	0.32*
[A]Session 18	0.36***	0.36*	0.40**
Asteroids 1	0.05	-0.08	0.08
Asteroids 2	0.13	-0.03	0.33*
Asteroids 3	0.24*	0.06	0.38**
Asteroids 4	0.07	-0.07	0.18
Tempest 1	-0.26*	-0.27	-0.35*
Tempest 2	0.02	-0.10	0.02
Tempest 3	-0.09	0.10	-0.20
Tempest 4	0.03	0.03	0.03

NOTE: [A]Keyboard Version of Space Fortress. [B]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). *$p < .05$; **$p < .01$; ***$p < .001$. All tests are two-tailed.

The results presented in Table 12 indicate that declarative knowledge was moderately related to Space Fortress performance across all sessions. For the total sample, the mean correlation for the acquisition phase were 0.33 (*SD* = 0.08). The mean correlations for dyads and individuals were 0.35 (*SD* = 0.09) and 0.31 (*SD* = 0.06) respectively. Similar effects were obtained for the re-acquisition phase (r = 0.33 [*SD* = 0.05]; r = 0.35 [*SD* = 0.05]; r = 0.31 [*SD* = 0.03] for the total sample, dyads, and individuals, respectively). These effects were generally larger than those obtained for cognitive ability.

Figure 13 presents the overall declarative knowledge-performance correlations and the declarative knowledge-performance correlations for each training protocol across all training sessions. Table 13 presents the results of hierarchical regression analyses to test for the stability of the declarative knowledge-performance relationship. No significant results were obtained for any of the models. Thus, the declarative knowledge-performance relationship appears to remain stable for both conditions throughout training.

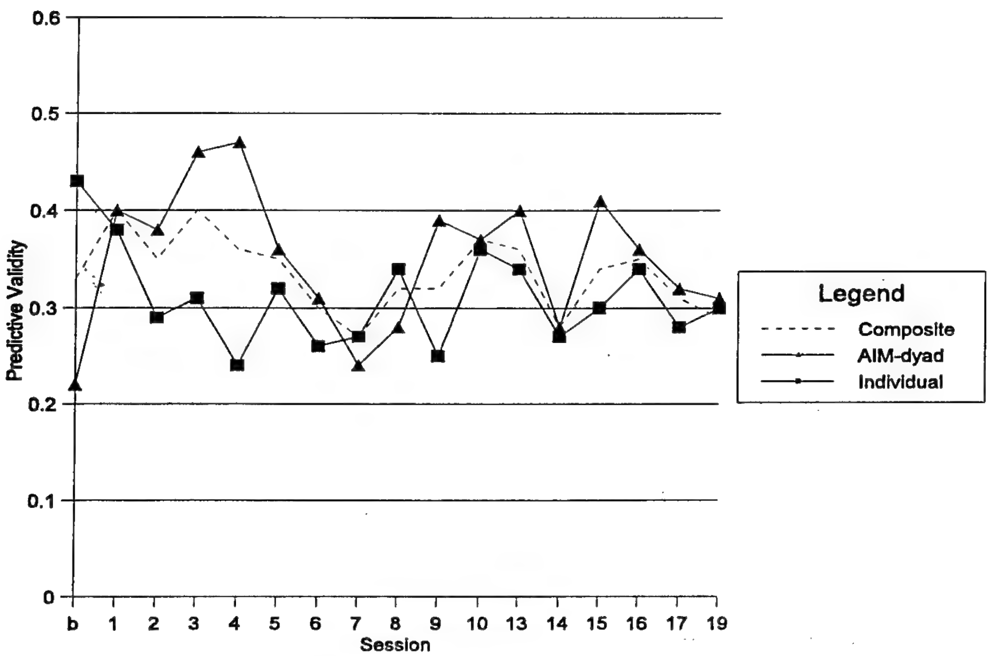

Figure 13. Declarative knowledge-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 13

Hierarchical Regression of Declarative Knowledge-Performance Correlations on Session and Condition to Test for Nature of fit

Models	^A^B	Model R^2	ΔR^2
Session	-0.0199	0.0265	
Session2	0.0010	0.0346	0.0081
Training Condition	-0.0503	0.1354	0.1008
Condition * Session	0.0260	0.1381	0.0027
(Condition * Session)^2	-0.0014	0.1921	0.0540

NOTE: AIM-dyad = 1; Individual-based = 0. AThe numbers presented are from the final regression equation. *p < .05; **p < .01; ***p < .001.

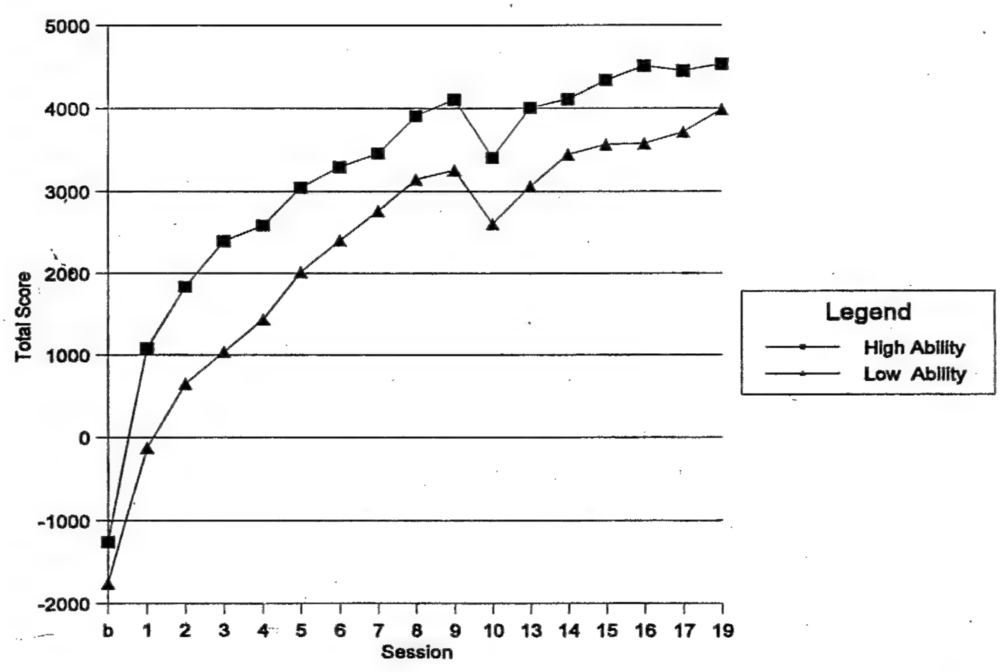

Figure 14. High and low declarative knowledge trainees' Space Fortress performance across training sessions.

To further examine the nature of the declarative knowledge-performance relationship, differences between high and low declarative knowledge trainees on Space Fortress sessions were assessed. To accomplish this, a median split was used to create high and low ability groups. Results of a 2x17 (declarative knowledge x practice) mixed factors ANOVA indicated that practice did not eliminate performance differences between high and low declarative knowledge

trainees (see Figure 14). The results of these analyses indicated that there was a significant effect for declarative knowledge (F [1, 87] = 8.67, p = 0.0041), a significant practice effect (F [16, 1392] = 371.20, p = 0.0001), and a significant declarative knowledge-by-practice interaction (F [16, 1392] = 2.27, p = 0.0028). Using the Bonferroni correction, paired t-tests indicate that trainees high in declarative knowledge scored particularly better on Space Fortress during sessions 1-4 compared to trainees with low declarative knowledge scores.

Psychomotor Ability. The objectives and analyses for this variable were identical to those previously presented for the other individual difference variables. The results presented in Table 14 indicate that psychomotor ability was moderately related to Space Fortress performance across all sessions. For the total sample, the mean correlation for the acquisition phase was 0.43 (SD = 0.03). The mean correlations for dyads and individuals were 0.48 (SD = 0.05) and 0.38 (SD = 0.05) respectively. Similar effects were obtained for the re-acquisition phase (r = 0.42 [SD = 0.03]; r = 0.44 [SD = 0.11]; r = 0.37 [SD = 0.06] for the total sample, dyads, and individuals, respectively). These effects were generally larger than those obtained for both cognitive ability and declarative knowledge.

Figure 15 presents the overall psychomotor ability-performance correlations and the psychomotor ability-performance correlations for each training protocol across all training sessions. Table 15 presents the results of hierarchical regression analyses to test for the stability of the psychomotor ability-performance relationship. The incremental effect for training condition was significant. This suggests that across all sessions, the relationship between psychomotor ability and Space Fortress performance was stronger for dyads compared to individuals. None of the other incremental effects were significant. Thus, it appears that psychomotor ability is a stable predictor of overall Space Fortress performance for both training protocols. However, it should be pointed out that the baseline difference (r = 0.24) between the dyads and individuals could explain the overall difference between the two protocols throughout training. Further analyses failed to obtain any differences between the two protocols on mean psychomotor ability, variability on either psychomotor ability or baseline scores, and the reliability of psychomotor ability measure. Therefore, it is likely that the overall difference between the two groups stems from a nonequivalence in predispositions to rely on psychomotor ability when playing Space Fortress.

Table 14

Correlations Between Psychomotor Ability and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	AIM-Dyad	Individual-Based
Session 0	0.39***	0.51***	0.27
Session 1	0.45***	0.47**	0.42**
Session 2	0.40***	0.46**	0.34*
Session 3	0.39***	0.42**	0.37**
Session 4	0.37***	0.43**	0.30*
Session 5	0.41***	0.41**	0.40**
Session 6	0.49***	0.53***	0.44**
Session 7	0.44***	0.47**	0.41**
Session 8	0.47***	0.54***	0.41**
Session 9	0.45***	0.57***	0.35*
Session 10	0.44***	0.52***	0.38**
Session 13	0.46***	0.51***	0.43**
Session 14	0.45***	0.46**	0.43**
Session 15	0.45***	0.53***	0.41**
Session 16	0.40***	0.44**	0.36*
Session 17	0.39***	0.50***	0.28*
Session 19	0.39***	0.52***	0.30*
[B]Skill Loss	0.06	0.10	0.01
TRANSFER TASKS			
[A]Session 12	0.36***	0.40*	0.35*
[A]Session 18	0.30**	0.38*	0.22
Asteroids 1	0.23*	0.29	0.20
Asteroids 2	0.27*	0.29	0.25
Asteroids 3	0.29**	0.31	0.27
Asteroids 4	0.24*	0.49**	0.02
Tempest 1	0.01	0.33*	-0.09
Tempest 2	-0.02	-0.03	-0.03
Tempest 3	0.08	0.14	0.07
Tempest 4	0.13	-0.02	0.23

NOTE: [A]Keyboard Version of Space Fortress. [B]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). **p* < .05; ***p* < .01; ****p* < .001. All tests are two-tailed.

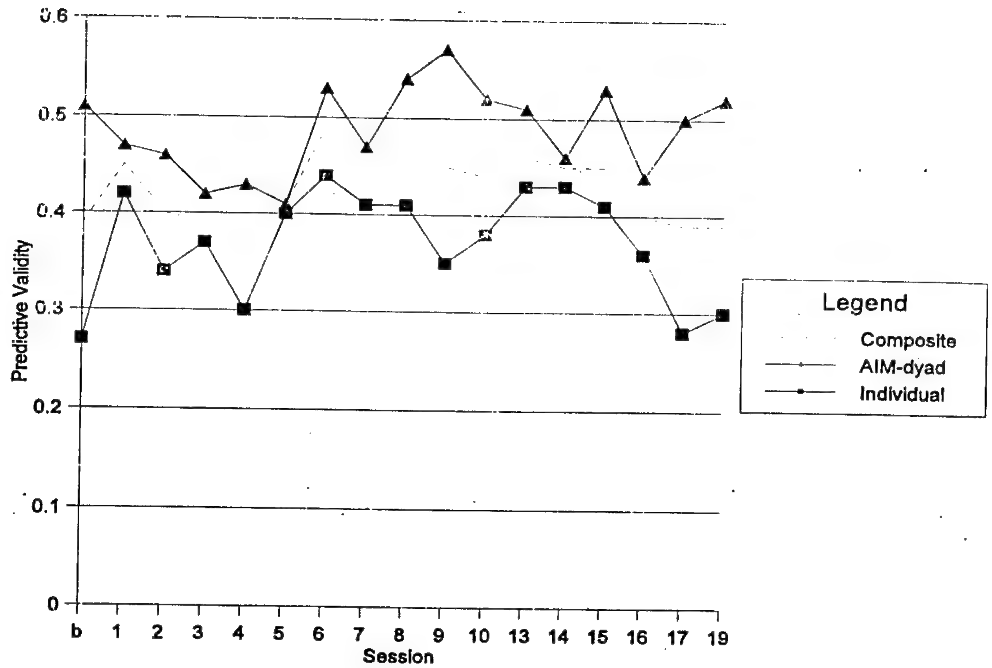

Figure 15. Psychomotor ability-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 15

Hierarchical Regression of Psychomotor Ability-Performance Correlations on Session and Condition to Test for Nature of fit

Models	${}^{a}B$	Model R^2	ΔR^2
Session	0.0291	0.0092	
Session^2	-0.0016	0.0733	0.0641
Training Condition	0.1707	0.6532***	0.5799***
Condition * Session	-0.0224	0.6621***	0.0089
(Condition * Session)^2	0.0014	0.7005***	0.0384

NOTE: AIM-dyad = 1; Individual-based = 0. [a]The numbers presented are from the final regression equation. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

To further examine the nature of the psychomotor ability-performance relationship, differences between high and low psychomotor ability trainees on Space Fortress sessions were assessed. To accomplish this, a median split was used to create high and low ability groups. Results of a 2x17 (psychomotor ability x practice) mixed factors ANOVA indicated that practice did not eliminate performance differences between high and low psychomotor ability trainees

(see Figure 16). The results of these analyses indicated that there was a significant effect for psychomotor ability (F [1, 87] = 16.56, p = 0.0041), and a significant practice effect (F [16, 1392] = 400.62, p = 0.0001). The psychomotor ability-by-practice interaction was not significant (F [16, 1392] = 1.56, p = 0.0727). Therefore, the results indicate that the rate of improvement on Space Fortress was the same for both high and low psychomotor ability trainees.

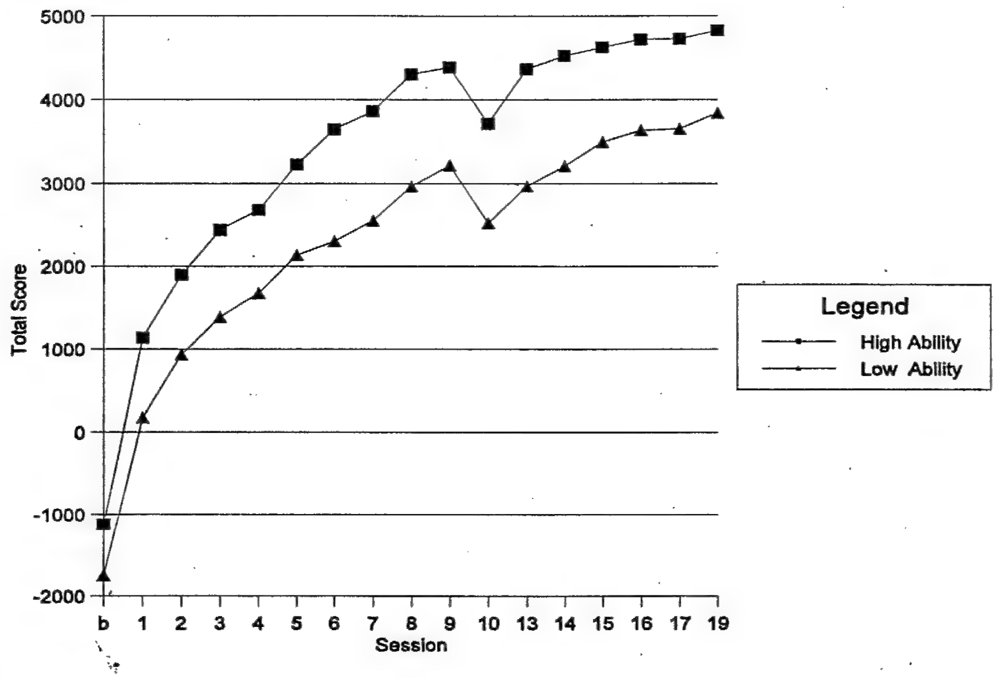

Figure 16. High and low psychomotor ability trainees' Space Fortress performance across training sessions.

Spatial Processing Speed. The objectives and analyses for this variable were identical to those previously presented for the other individual difference variables. The results presented in Table 16 indicate that the relationships between spatial processing speed and Space Fortress performance were very weak. For the total sample, the mean correlation for the acquisition phase was 0.15 (SD = 0.12). The mean correlations for dyads and individuals were 0.24 (SD = 0.09) and 0.06 (SD = 0.05) respectively. Similar effects were obtained for the re-acquisition phase (r = 0.13 [SD = 0.05]; r = 0.15 [SD = 0.07]; r = 0.11 [SD = 0.02] for the total sample, dyads, and individuals, respectively). These effects were much lower than those obtained for cognitive ability, declarative knowledge, and psychomotor ability.

Table 16

Correlations Between Spatial Processing Speed and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	AIM-Dyad	Individual-based
Session 0	0.19	0.33*	0.10
Session 1	0.14	0.32*	0.02
Session 2	0.11	0.32*	-0.02
Session 3	0.15	0.30*	0.08
Session 4	0.15	0.28	0.08
Session 5	0.12	0.19	0.09
Session 6	0.03	0.13	-0.01
Session 7	0.10	0.09	0.12
Session 8	0.08	0.18	0.04
Session 9	0.16	0.27	0.11
Session 10	0.13	0.12	0.15
Session 13	0.14	0.21	0.10
Session 14	0.11	0.15	0.09
Session 15	0.13	0.25	0.08
Session 16	0.12	0.16	0.11
Session 17	0.09	0.09	0.11
Session 19	0.12	0.07	0.14
[B]Skill Loss	0.06	0.27	-0.07
TRANSFER TASKS			
[A]Session 12	0.24*	0.05	0.34*
[A]Session 18	0.27**	0.23	0.30*
Asteroids 1	-0.03	0.26	-0.14
Asteroids 2	0.09	0.20	-0.00
Asteroids 3	0.12	0.21	0.07
Asteroids 4	-0.06	0.02	-0.11
Tempest 1	-0.13	0.15	-0.18
Tempest 2	0.07	0.08	0.09
Tempest 3	-0.01	0.10	-0.02
Tempest 4	0.03	0.14	-0.03

NOTE: [A]Keyboard Version of Space Fortress. [B]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). **p* < .05; ***p* < .01; ****p* < .001. All tests are two-tailed.

Figure 17 presents the overall spatial processing speed-performance correlations and the spatial processing speed-performance correlations for each training protocol across all training sessions. Table 17 presents the results of hierarchical regression analyses to test for the stability of the relationships presented in Figure 17. Significant incremental effects were obtained for both training condition and the condition-by-session interaction. These results indicate that across all sessions, the relationship between spatial processing speed and Space Fortress performance was stronger for dyads compared to individuals. Furthermore, the spatial processing speed-performance relationship appears to have decreased over training sessions for the dyad protocol; whereas the spatial processing speed-performance relationship was stable for the individual protocol. However, it should be pointed out that the baseline difference (r = 0.23) between the dyads and individuals could explain both the overall difference between the two protocols throughout training and the decreasing trend in the dyad protocol. Further analyses failed to obtain any differences between the two protocols on mean spatial processing speed, variability on either spatial processing speed or baseline scores, and the reliability of spatial processing speed measure. Therefore, it is likely that the overall difference between the two groups stems from a nonequivalence in predispositions to rely on spatial processing speed when playing Space Fortress, and the decreasing trend in the dyad protocol could very well be the result of a spuriously high spatial processing speed-performance relationship early in training for dyad trainees. Conversely, it could be argued that the differences between the two protocols are due to a spuriously weak spatial processing speed-performance relationship in the individual condition.

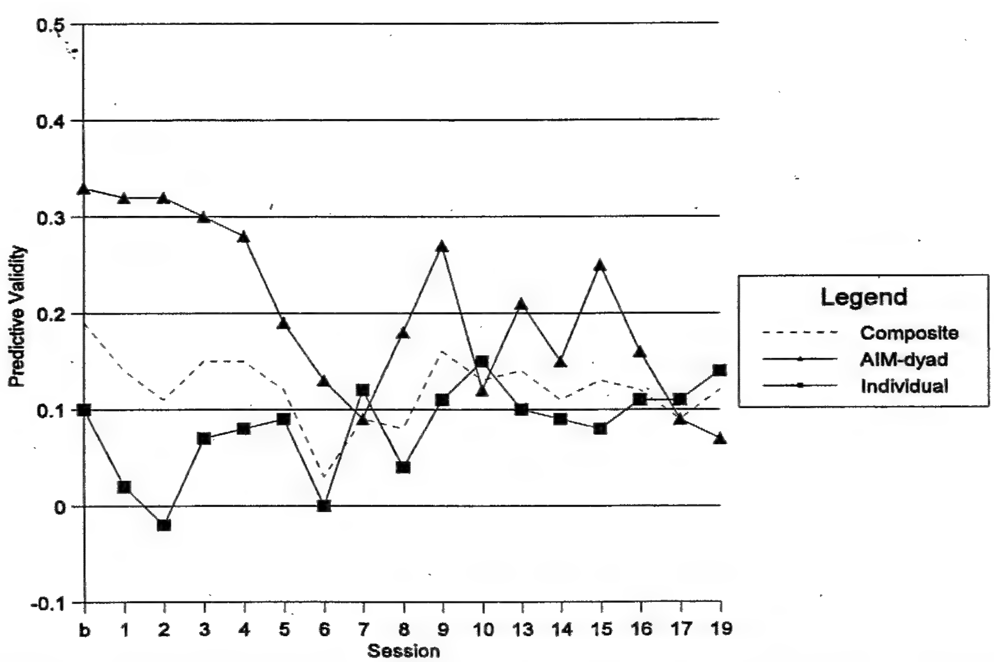

Figure 17. Spatial processing speed-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 17

Hierarchical Regression of Spatial Processing Speed-Performance Correlations on Session and Condition to Test for Nature of fit

Models	^B	Model R^2	ΔR^2
Session	0.0032	0.0420	
$Session^2$	0.0001	0.0511	0.0091
Training condition	0.3179	0.4853***	0.4342**
Condition * Session	-0.0289	0.7163***	0.2310**
$(Condition * Session)^2$	0.0006	0.7212***	0.0049

Note: AIM-dyad=1; Individual-based=0. ^The numbers presented are from the final regression equation. *p<.05; **p<.01; ***p<.001.

To further examine the nature of the spatial processing speed and performance on Space Fortress, differences between high and low spatial processing speed trainees on Space Fortress sessions were assessed. To accomplish this, a median split was used to create high and low ability groups. Results of a 2x17 (spatial processing speed x practice) mixed factors ANOVA indicated that there were performance differences between high and low ability trainees (see Figure 18) across the Space Fortress sessions. Specifically, the results of these analyses indicated that the effect for spatial processing speed was not significant ($F[1, 87] = 0.32$, $p = 0.5741$). Neither was the ability-by-practice interaction ($F[16, 1392] = 0.46$, $p = 0.9664$). On the other hand, the practice effect was significant ($F[16, 1392] = 396.11$; $p = 0.0001$) indicating that although both high and low ability trainees improved with performance, they did so at the same rate.

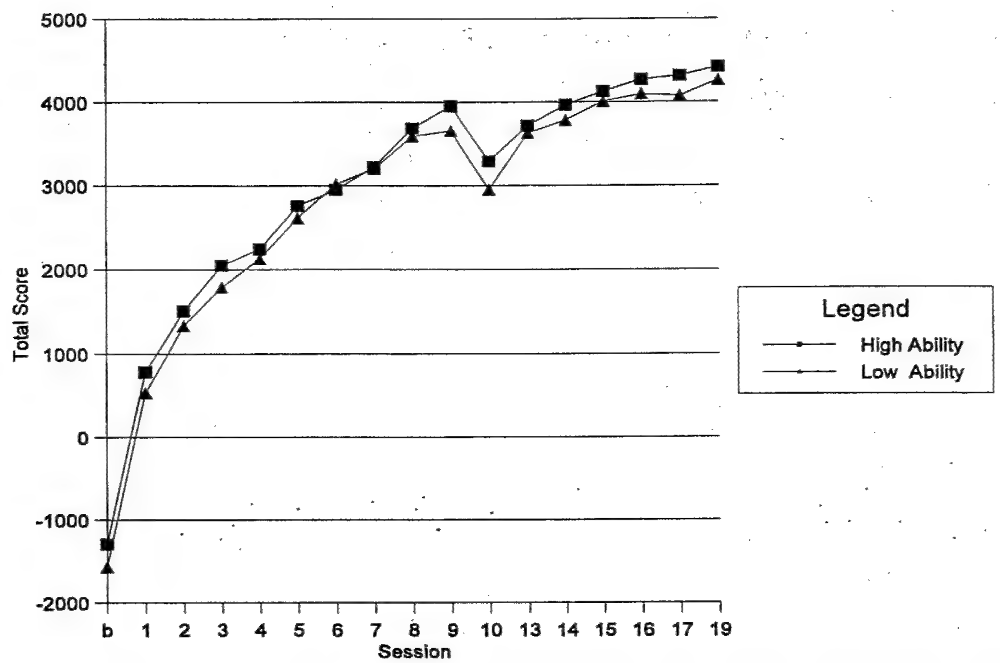

Figure 18. High and low spatial processing speed trainees' Space Fortress performance across training sessions.

Spatial Working Memory. The objectives and analyses for this variable were identical to those previously presented for the other individual difference variables. The results presented in Table 18 indicate that the overall relationships between spatial working memory and Space Fortress performance were relatively weak. For the total sample, the mean correlations between spatial working memory and performance were 0.22 (*SD* = 0.16), and 0.21 (*SD* = 0.13) for the acquisition and re-acquisition phases, respectively.

However, the effects were much stronger for dyads than individuals. For the acquisition and re-acquisition phases, the mean correlations for dyads were 0.36 (*SD* = 0.05) and 0.33 (*SD* = 0.07), compared to 0.08 (*SD* = 0.05) and 0.09 (*SD* = 0.03) for individuals.

Table 18

Correlations Between Spatial Working Memory and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	AIM-Dyad	Individual-based
Session 0	0.20	0.38*	0.02
Session 1	0.26*	0.40*	0.13
Session 2	0.27*	0.49**	0.06
Session 3	0.22*	0.36*	0.10
Session 4	0.19	0.37*	0.03
Session 5	0.24*	0.39*	0.14
Session 6	0.18	0.34*	0.06
Session 7	0.16	0.29	0.07
Session 8	0.22*	0.34*	0.14
Session 9	0.16	0.33*	0.03
Session 10	0.24*	0.36*	0.16
Session 13	0.21*	0.32*	0.13
Session 14	0.20	0.35*	0.06
Session 15	0.17	0.37*	0.05
Session 16	0.23*	0.41**	0.08
Session 17	0.15	0.22	0.10
Session 19	0.18*	0.28	0.10
[B]Skill Loss	-0.14	-0.03	-0.22
TRANSFER TASKS			
[A]Session 12	0.22*	0.29	0.13
[A]Session 18	0.20	0.27	0.13
Asteroids 1	0.01	0.21	-0.09
Asteroids 2	0.08	0.11	0.05
Asteroids 3	-0.01	-0.06	0.04
Asteroids 4	0.22*	0.32*	0.14
Tempest 1	-0.09	0.07	-0.14
Tempest 2	-0.10	-0.15	-0.07
Tempest 3	0.04	0.07	0.05
Tempest 4	0.01	-0.01	0.04

NOTE: [A]Keyboard Version of Space Fortress. [B]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). *p < .05; **p < .01; ***p < .001. All tests are two-tailed.

Figure 19 presents the overall spatial working memory-performance correlations and the spatial working memory-performance correlations for each training protocol across all training sessions. Table 19 presents the results of hierarchical regression analyses to test for the stability of the relationships presented in Figure 19. Significant incremental effects were obtained for both training condition and the condition-by-session interaction. These results indicate that across all sessions, the relationship between spatial working memory and Space Fortress performance was stronger for dyads compared to individuals. Furthermore, the spatial working memory-performance relationship appears to have decreased over training sessions for the dyad protocol; whereas the spatial working memory-performance relationship was stable for the individual protocol. However, it should be pointed out that the baseline difference (r = 0.36) between the dyads and individuals could explain both the overall difference between the two protocols throughout training and the decreasing trend in the dyad protocol. Further analyses failed to obtain any differences between the two protocols on mean spatial working memory, variability on either spatial working memory or baseline scores, and the reliability of spatial working memory measure. Therefore, it is likely that the overall difference between the two groups stems from a nonequivalence in predispositions to rely on spatial working memory when playing Space Fortress, and the decreasing trend in the dyad protocol could very well be the result of a spuriously high spatial processing speed-performance relationship early in training for dyad trainees. Conversely, it could be argued that the differences between the two protocols are due to a spuriously weak spatial working memory-performance relationship in the individual condition.

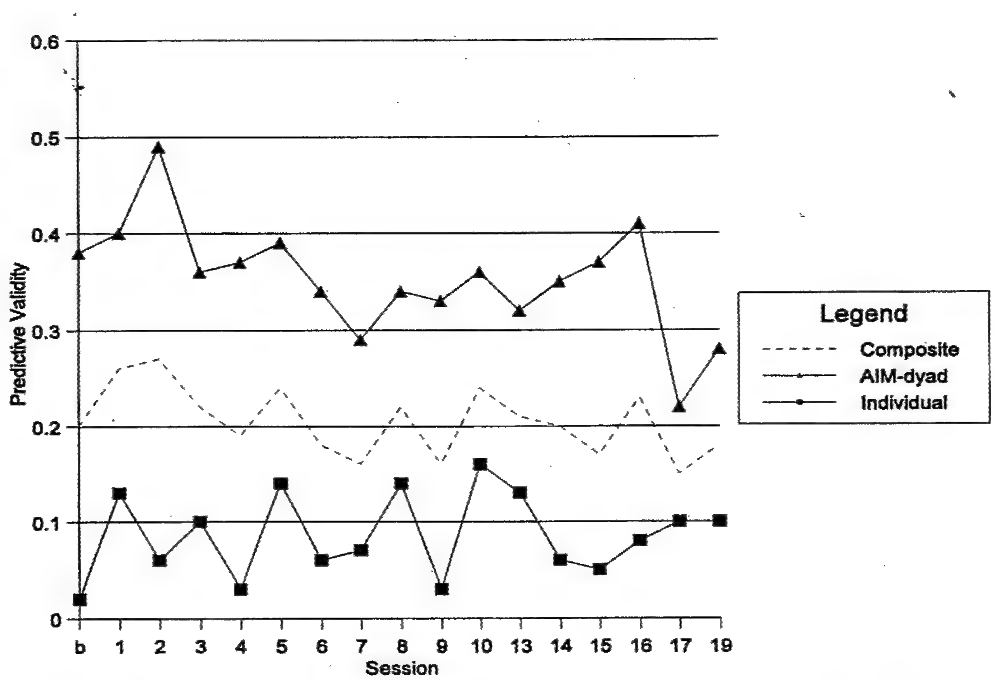

Figure 19. Spatial working memory-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 19

Hierarchical Regression of Spatial Working Memory-Performance Correlations on Session and Condition to Test for Nature of fit

Models	[A]B	Model R^2	ΔR^2
Session	0.0070	0.0082	
$Session^2$	-0.0003	0.0085	0.0003
Training	0.3612	0.8843***	0.8758***
Condition * Session	-0.0150	0.9030***	0.0187*
$(Condition * Session)^2$	0.0004	0.9036***	0.0009

Note : AIM-dyad=1; Individual-based=0. [A]The numbers presented are from the final regression equation. *$p < .05$; **$p < .01$; ***$p < .001$.

To further examine the nature of the relationship between spatial working memory and performance on Space Fortress, differences between high and low spatial working memory trainees on Space Fortress sessions were assessed. To accomplish this, a median split was used to create high and low ability groups. Results of a 2x17 (spatial working memory x practice) mixed factors ANOVA indicated that there were no performance differences between high and low ability trainees (see Figure 20) across the Space Fortress sessions. Specifically, the results of these analyses indicated that the effect for spatial working memory was not significant (F [1, 87] = 2.01, p = 0.1594). Neither was the ability-by-practice interaction (F[16, 1392] = 0.61, p = 0.8791). On the other hand, the practice effect was significant (F[16, 1392] = 396.86, p = 0.0001) indicating that although both high and low ability trainees improved with practice, they did so at the same rate.

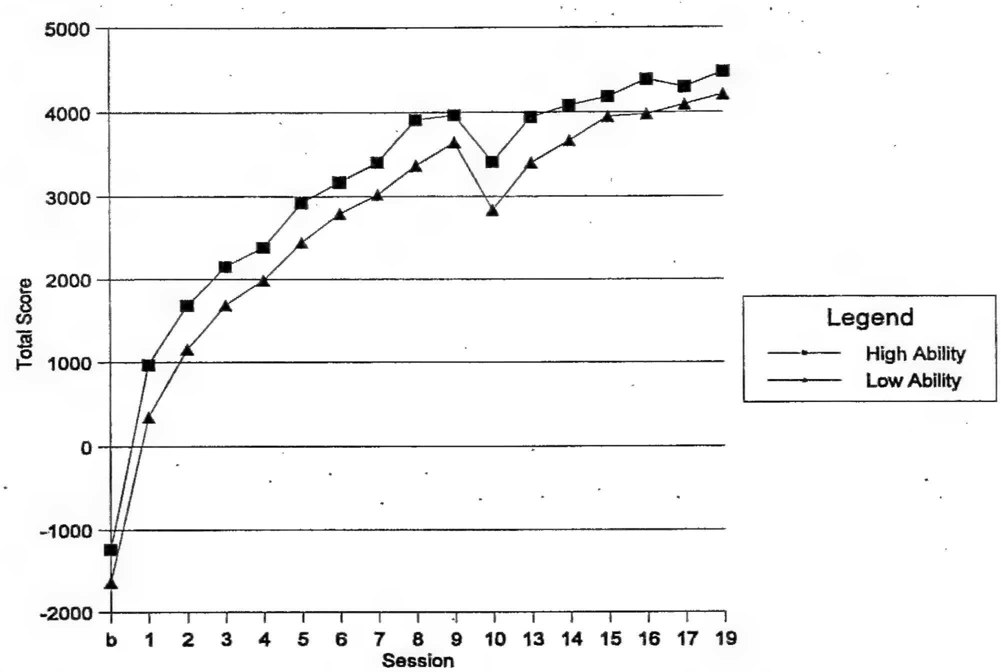

Figure 20. High and low spatial working memory trainees' Space Fortress performance across training sessions.

Visual Attention. The objectives and analyses for this variable were identical to those previously presented for the other individual difference variables. The results presented in Table 20 indicate that the overall relationships between visual attention and Space Fortress performance were moderate. For the total sample, the mean correlations between visual attention and performance were 0.34 (SD = 0.06), and 0.30 (SD = 0.03) for the acquisition and re-acquisition phases, respectively.

Similar effects were obtained for dyads and individuals. For the acquisition and re-acquisition phases, the mean correlations for dyads were 0.33 (SD = 0.06) and 0.30 (SD = 0.04), compared to 0.36 (SD = 0.05) and 0.30 (SD = 0.02) for individuals.

Table 20

Correlations Between Visual Attention and Space Fortress Performance (and Transfer Tasks)

Space Fortress Session	Total Sample	AIM-Dyad	Individual-based
Session 0	0.30**	0.26	0.34*
Session 1	0.31**	0.25	0.39**
Session 2	0.40***	0.39*	0.42**
Session 3	0.36***	0.29	0.45**
Session 4	0.32**	0.29	0.37**
Session 5	0.37***	0.36*	0.40**
Session 6	0.32**	0.32*	0.33*
Session 7	0.34**	0.39*	0.28*
Session 8	0.30**	0.30	0.31*
Session 9	0.36***	0.41**	0.32*
Session 10	0.32**	0.34*	0.30*
Session 13	0.28**	0.28	0.29*
Session 14	0.30**	0.29	0.31*
Session 15	0.30**	0.35*	0.29*
Session 16	0.33**	0.35*	0.32*
Session 17	0.26**	0.24	0.29*
Session 19	0.29**	0.28	0.32*
[B]Skill Loss	0.10	0.12	0.08
TRANSFER TASKS			
[A]Session 12	0.29**	0.28	0.32*
[A]Session 18	0.21	0.21	0.22
Asteroids 1	0.07	0.12	0.04
Asteroids 2	0.09	0.02	0.18
Asteroids 3	-0.01	-0.20	0.19
Asteroids 4	0.14	0.07	0.22
Tempest 1	0.02	0.13	-0.02
Tempest 2	-0.02	-0.23	0.09
Tempest 3	0.01	0.00	0.01
Tempest 4	-0.03	-0.07	0.00

NOTE: [A]Keyboard Version of Space Fortress. [B]Skill loss was operationalized as the performance difference between the last acquisition trial (Session 9) and the first session after the eight-week non-practice retention interval (Session 10). *$p < .05$; **$p < .01$; ***$p < .001$. All tests are two-tailed. All correlations have been transposed so that a positive correlation indicates better performance on both measures.

Figure 21 presents the overall visual attention-performance correlations and the visual attention-performance correlations for each training protocol across all training sessions. Table 21 presents the results of hierarchical regression analyses to test for the stability of the relationships presented in Figure 21. Significant incremental effects were obtained for both the condition-by-session and condition-by-session squared interactions. These results indicate that across all sessions, the relationship between visual attention and performance was curvilinear for dyads, such that the form of the relationship was that of an inverted U where the relationship was strongest mid-way through training. For individuals, the magnitude of the visual attention-performance relationship decreased over training sessions.

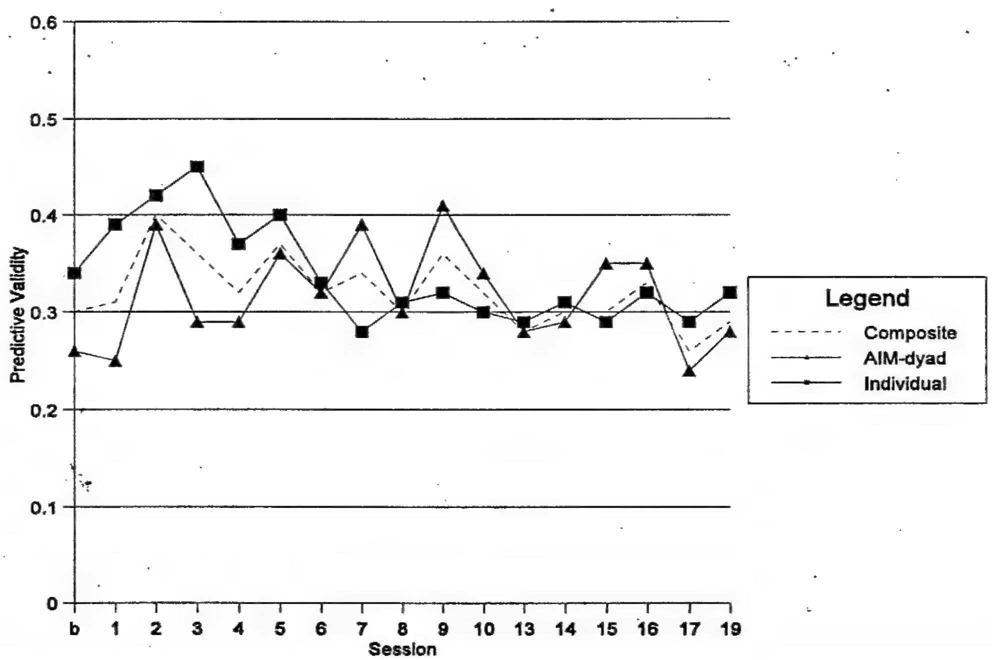

Figure 21. Visual attention-Space Fortress performance correlations for AIM-dyad and individual conditions across training sessions.

Table 21

Hierarchical Regression of Visual Attention-Performance Correlations on Session and Condition to Test for Nature of fit

Models	^A^*B*	Model R^2	ΔR^2
Session	-0.0149	0.1164*	
Session^2	0.0004	0.1456	0.0292
Training Condition	-0.1782	0.1845	0.0389
Condition * Session	0.0372	0.2961*	0.1116*
$(\text{Condition * Session})^2$	-0.0016	0.4221**	0.1260*

NOTE: AIM-dyad = 1; Individual-based = 0. [A]The numbers presented are from the final regression equation. **p* < .05; ***p* < .01; ****p* < .001.

To further examine the nature of relationship between visual attention and performance on Space Fortress, differences between high and low visual attention trainees on Space Fortress sessions were assessed. To accomplish this, a median split was used to create high and low ability groups. Results of a 2x17 (visual attention x practice) mixed factors ANOVA indicated that there were performance differences between high and low ability trainees (see Figure 22) across the Space Fortress sessions. Specifically, significant effects were obtained for visual attention (F[1, 87] = 7.24, p = 0.0086), and practice (F[16, 1392] = 397.60, p = 0.0001). However, the attention-by-practice interaction was not significant (F[16, 1392] = 0.78, p = 0.7111). Thus although both high and low ability trainees improved with performance, they did so at the same rate.

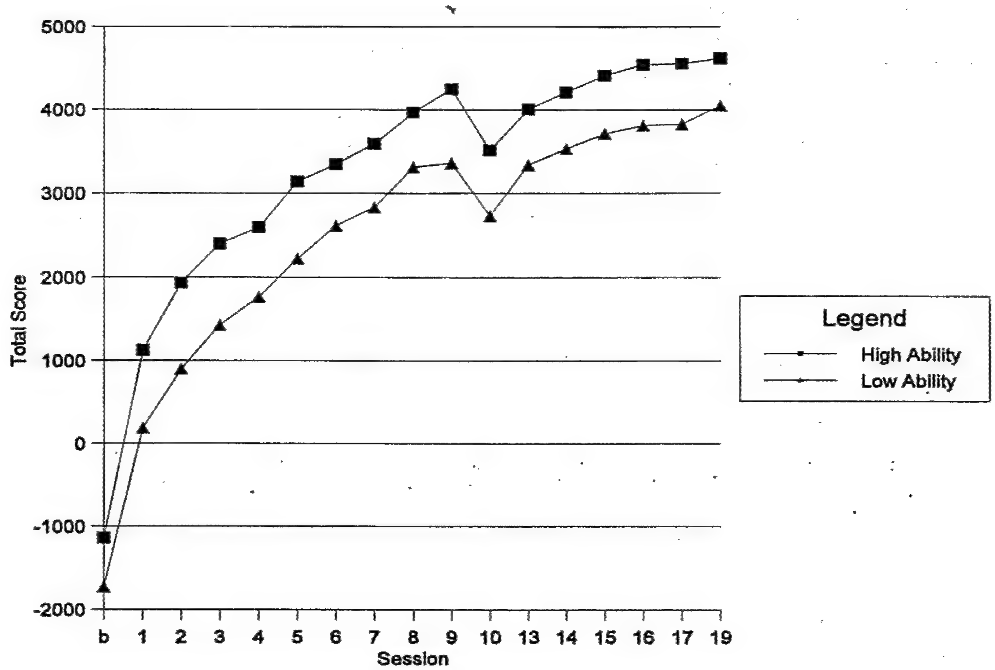

Figure 22. High and low visual attention trainees' Space Fortress performance across training sessions.

DISCUSSION

The first objective of the present study was to evaluate the comparative effectiveness of the AIM-dyad and individual-based protocols in the context of complex skill acquisition, loss, and re-acquisition. The results obtained here are consistent with previous research (e.g., Arthur et al., 1995; Shebilske et al., 1992) and provide additional support for the efficiency gains associated with the AIM-dyad protocol. In spite of having half the amount of hands-on practice and simultaneously training two individuals at the same computer station, participants in the AIM-dyad achieved the same level of task performance as those in the individual-based protocol. Furthermore, over an 8-week non-practice interval, participants' level of skill loss and rate of skill re-acquisition were not different from participants in the individual-based protocol. In the parlance of significance testing, this finding is based on a failure to reject the null hypothesis. Given some of the problems associated with significance testing (see Cohen, 1994; Schmidt, 1996), the magnitude of the performance difference in d's was also computed and reported in Table 4.

Differences between the AIM-dyad and individual-based protocols in ship control strategy in the skill acquisition and re-acquisition phases were also investigated. As expected, the results indicated that in general, trainees who had more effective control strategies had higher Space Fortress scores. Overall, the two protocols did not differ on ship control strategies during either the acquisition or re-acquisition phases. However, although dyads had less effective strategies during the early sessions of skill acquisition, they surpassed individuals in the latter sessions (see Figure 4). Finally, the two protocols did not differ in terms of strategy loss.

In general, the results of the present study demonstrate the robustness of the efficiency gains associated with the AIM-dyad protocol by showing that the 100% increase in the effective use of time and resources is still present after an extended period of non-practice. The results of the present study also suggest that the social and cognitive variables aligned to favor the AIM protocol and to compensate for the reduction in hands-on practice (Arthur et al., 1995; Jordan, 1991) are effective in not only facilitating the skill acquisition of AIM-dyad trainees, but they also allow them to achieve and maintain the same level of performance as trainees in the individual-based protocol. As noted by Shebilske et al. (1992), it also suggests that the AIM-dyad protocol could substantially reduce training costs in military and industrial applications by doubling the number of personnel trained on automated instructional systems with no increase in time, software, or hardware costs (Shebilske et al., in press).

The second objective of this project was to investigate the effect and role of retention in the transfer of acquired skills. The intent was to further assess the comparative effectiveness of the protocols in terms of the facilitation or inhibition of the transfer of acquired skills. This research objective was primarily exploratory in nature and was intended to serve as a pilot run for a more rigorous investigation of this issue. Consequently, although their appropriateness as transfer tasks could be questioned, the two tasks that were used were Asteroids, which was considered to be a positive transfer task, and Tempest which was considered to be a neutral transfer task. The keyboard version of

Space Fortress was also investigated as a positive transfer task. The results of the transfer analyses indicated that the two protocols did not differ on any of the transfer tasks. This finding should, however, be cautiously interpreted since, as previously noted, there is some question about the suitability of Asteroids and Tempest as transfer tasks. Future research is planned to further investigate this issue using more appropriate tasks, namely Phoenix (CITE) and Loader (CITE).

The third research objective was to assess the ability of specified individual difference variables to predict not only skill acquisition in original learning, but also the amount of skill loss and re-acquisition as well. The stability of ability-performance relationships were also investigated. Of the 6 individual difference variables, the strongest predictor was psychomotor ability (r = 0.43 for the total sample), followed by visual attention (r = 0.34 for the total sample). The weakest predictor was spatial processing speed (r = 0.15). The rate of improvement for high and low ability trainees on all predictors were also the same. Stable ability-performance relationships were obtained for only declarative knowledge and psychomotor ability. Finally, the rate of acquisition and re-acquisition was the same for high and low ability trainees on all the predictors.

Implications, Limitations, and Suggestions for Future Research

First, in terms of long-term retention, there are a number of plausible explanations for the similar performance levels obtained for the AIM-dyad and individual-based protocols. Although not demonstrated in the present study, analysis of a larger data set by Day et al. (in press) led them to conclude that because of the substantial difference in the amount of hands-on practice between the two protocols, they might be differentially utilizing trainees' abilities. Specifically, due to the significant reduction in hands-on practice and emphasis placed on observational learning, trainees in the AIM protocol may rely more heavily on cognitive ability compared to those in the individual-based protocol. Conversely, trainees in the individual-based protocol may rely more heavily on psychomotor ability. In contrast to their study, differential effects were obtained for spatial processing speed, spatial working memory, and visual attention, but not for cognitive and psychomotor ability in the present study. Consequently, the general suggestion is for future dynamic criteria or stability of ability-performance relationships research to address or recognize the role that the type of skill acquisition paradigm or training protocol may play in ability-performance relationships. With respect to the considerable attention aptitude-treatment interactions have received in the instructional and training literature (Ackerman, Sternberg, & Glaser, 1989; Cronbach & Snow, 1977; Pintrich, Cross, Kozma, & Mckeachie, 1986; Snow, 1989; Tannenbaum & Yukl, 1992) the importance of training protocol seems pertinent to discussions of ability-performance relationships.

The second issue is in reference to the measures used. Future research might consider alternative measures with better psychometric properties. For instance, the original Raven's Advanced Progressive Matrices (Raven et al., 1993) or even the short form (Arthur & Day, 1994) could be used instead of the computer-administered analogue used here.

Also, the declarative knowledge test used was a recognition test. It could be reasonable argued that a recall test would have been more appropriate. Third, although participants were randomly assigned to conditions, there were substantial difference across protocols on some of the individual difference variables. Thus future work might consider matching participants on the basis of their predictor scores.

Fourth, the final sample size was relatively small. This was due to a relatively large attrition rate (i.e., 100 of the 189 individuals recruited dropped out of the study) arising primarily from the 8-week non-practice interval. Although there were no differences on any of the individual difference variables between the final sample and those who dropped out of the study, the problems associated with small sample sizes is well documented (e.g., Schmidt & Hunter, 1978). Furthermore, since attrition from extended training programs is an issue of interest in real life contexts, future research is planned to specifically investigate factors (such as work ethic [Woher & Miller, 1997]) that might predict participants attrition from the study.

Fifth, although results of the post-hoc analyses indicate that trainees may have approached mastery on the task, future research might consider training participants to a specified level of mastery instead of training for a fixed length of time. On a related note, the present study could be extended by using longer retention intervals. Specifically, although the 8-week non-practice interval is longer than that used in most of the extant literature (cf. Arthur et al., in press, Table 3), it would be interesting to see if the present results would be obtained with substantially longer non-practice intervals.

Sixth, the aiming task was used in the present study as both a screening tool and predictor. The magnitude of this potential problem is primarily a function of how many individuals were eliminated on the basis of their aiming task performance. Since only two individuals out of the total number who were recruited failed the aiming task, using it for both purposes did not make much, if any difference, in terms of restricting the data.

Seventh, although "transfer" has many meanings, it was used in the present study to mean the savings in learning on one task (transfer task - i.e., Asteroids, Tempest, or keyboard version of Space Fortress) due to earlier training on a different task (training task - i.e., Space Fortress). This was assessed by testing for differences on the transfer tasks between the two protocols. Although the present design did not lend itself to their use, there are alternative operationalizations of transfer (Singley & Anderson, 1989; VanLehn, 1996). For instance, transfer can be expressed as a ratio -- specifically, the time saved in learning the transfer task divided by the time spent learning the training task. The intent is to use this paradigm in planned, more rigorous investigations of this issue.

Finally, the influence of individual differences on learning and performance in group settings has appeared repeatedly in the group dynamics literature (Driskell, Hogan, & Salas, 1987). Along these lines, in an investigation of the role of

interaction anxiety (Leary, 1983) on the effectiveness of the AIM-dyad protocol, Arthur et al. (1996), obtained a significant interaction between training protocol (AIM versus individual) and trainees' level of interaction anxiety. This research suggests that individual differences may play a role in not only the assignment into training protocols, but also the pairing of trainees. Specifically, pairing strategies can be important in the training of dyads and crews because there can be instances where the dyadic or group composition can be detrimental (Brooks, Ebner, Manning, & Balson, 1985), beneficial, or at the very least, have no negative effects (Dossett & Hulvershorn, 1983) on individual trainee performance. In an extension of Arthur et al. (1996), Tubre (1997) presents preliminary evidence which suggests that in the AIM-dyad protocol, having a partner who is high in interaction anxiety appears to be somewhat of an advantage for a trainee who is high in interaction anxiety, whereas a low-interaction-anxiety trainee seems to benefit from having a low-interaction-anxiety partner. Thus, the utility of individual differences in determining both the assignment to, and optimal pairing strategies within the AIM protocol warrants future research.

In conclusion, the present findings, which show that the efficiency gains associated with the AIM-dyad protocol are still present within the context of skill retention and re-acquisition and, therefore, demonstrate the robustness of these efficiency gains, provide strong support and justification for the ongoing use of innovative dyadic protocols for the training of pilots and navigators in both military and non-military settings (Johnston et al., 1994; Shebilske et al., in press) - examples of situations in which trainees may be subjected to extended periods of non-practice.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive activities and information processing. *Journal of Experimental Psychology-General, 117,* 288-318.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77,* 598-614.

Ackerman, P. L., Sternberg, R. J., & Glaser, R. (Eds.). *Learning and individual differences: Advances in theory and research.* New York: W. H. Freeman.

Adams, J. A., & Hufford, L. E. (1962). Contributions of a part-task trainer to the learning and relearning of a time-shared flight maneuver. *Human Factors, 14,* 295-308.

Alliger, G. M., Bennett, W., Jr, & Tannenbaum, S. I. (1995). Transfer of training: Comparison of paradigms. In K. Smith and E. Salas (Chairs), *The transfer of training problem: A multidisciplinary approach.* Symposium presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Arthur, W., Jr. (1991). *Individual differences in the prediction and training of complex perceptual-motor skill tasks: The development and validation of the Computer-Administered Visual Attention Test* (Tech. Report No. 8). College Station, TX: Texas A&M University, Psychology Department.

Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (in press). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance.*

Arthur, W., Jr., Day, D. V., (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54,* 394-403.

Arthur, W., Jr., Strong, M. H., Jordan, J. A., Williamson, J. E., Shebilske, W. L., & Regian. (1955). Visual attention: Individual differences in training and predicting complex task performance. *Acta Psychologica, 88,* 3-23.

Arthur, W., Jr., Strong, M. H., & Williamson, J. E. (1994). Validation of a visual attention test as a predictor of driving accident involvement. *Journal of Occupational and Organizational Psychology, 67,* 173-182.

Arthur, W., Jr., Young, B., Jordan, J. A., & Shebilske, W. (1996). Effectiveness of dyadic training protocols: The influence of trainee interaction anxiety. *Human Factors, 38,* 79-86.

Austin, J. T., Humphreys, L. G., & Hulin, C. L. (1989). Another review of dynamic criteria: A critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology, 42,* 597-612.

Avolio, B. J., Alexander, R. A., Barrett, G. V., & Sterns, H. L. (1981). Designing a measure of visual selective attention to assess individual differences in information processing. *Applied Psychological Measurement, 5,* 29-42.

Barrett, G. V., Alexander, R. A., & Doverspike, D. (1992). The implications for personnel selection of apparent declines in predictive validities over time: A critique of Hulin, Henry, and Noon. *Personnel Psychology, 45,* 601-617.

Barrett, G. V., Morris, S. B., & Alexander, R. A. (1993). The effect of factor analytic method on the interpretation of the changing abilities requirements model: An example of the tools-to-theory hypothesis. *Human Performance, 6*, 23-47.

Brooks, F. R., Ebner, D. G., Manning, D. T., & Balson, P. M. (1985). Influence of academic ability on partners' mastery of a stressful motor task. *Journal of Instructional Psychology, 12*, 19-23.

Carron, A. V. (1971). Effect of ability upon retention of a balance skill after two years. *Perceptual and Motor Skills, 33*, 527-529.

Carron, A. V., & Marteniuk, R. C. (1970). Retention as a balance skill as a function of initial ability level. *Research Quarterly, 41*, 478-483.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Day, E. A., Arthur, W., Jr., & Shebilske, W. L. (in press). Ability determinants of complex skill acquisition: Effects of training protocol. *Acta Psychologica*.

Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology, 43*, 717-744.

Donchin, E. (1989). The learning strategies project: Introductory remarks. *Acta Psychologica, 71*, 1-15.

Dossett, D. L., & Hulvershorn, P. (1983). Increasing technical training efficiency: Peer training via computer-assisted instruction. Journal of Applied Psychology, 68, 552-558.

Driskell, J. E., Hogan, R., & Salas, E. (1987). Personality and group performance. In C. Hendrick (Ed.), *Review of personality and social psychology* (Vol. 9, pp. 91-112). Newbury Park, CA: Sage.

Farr, M. J. (1987). *The long-term retention of knowledge and skills: A cognitive and instructional perspective*. New York: Springer-Verlag.

Ford, J. K., Quiñones, M., Sego, D., & Speer Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-527.

Foss, M. A., Fabiani, M., Mane, A. M., & Donchin, E. (1989). Unsupervised practice: The performance of the control group. *Acta Psychologica, 71*, 23-51.

Fox, W. L., Taylor, J. E., & Caylor, J. S. (1969). *Aptitude level and the acquisition of skills and knowledges in a variety of military training tasks* (HumRRO Tech. Rep. 69-6). Alexandria, VA: Human Resources Research Organization.

Frederiksen, J. R. & White, B. Y. (1989). An approach to training based on principled task decomposition. *Acta Psychologica, 71*, 89-146.

Goldberg, S. L., Drillings, M., & Dressel, J. D. (1981). *Mastery training: Effect on skill retention* (Technical Report 513). Alexandria, VA: U.S. Army Research Institute.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer and S. Korenblom (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*. (pp. 299-322).

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skill*, 33, 1335-1342.

Gopher, D., Weil, M., & Bareket, T. (1994). The transfer of skill from a computer game trainer to actual flight. *Human Factors, 36*, 387-405.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147-177.

Gottesfeld, N., & Arthur, W., Jr. (1994). *Openness and conscientiousness as predictors of performance on a complex perceptual-motor skill task*. Unpublished manuscript. Department of Psychology, Texas A&M University, College Station, TX.

Grimsley, D. L. (1969). *Acquisition, retention, and retraining: Training category IV personnel with low fidelity devices* (HumRRO Tech. Rep. 69-12). Alexandria, VA: Human Resources Research Organization.

Hagman, J. D. (1980a). *Effects of training schedule and equipment variety on retention and transfer of maintenance skill* (Research Report 1309). Alexandria, VA: Army Research Institute.

Hagman, J. D. (1980b). *Effects of training task repetition in retention and transfer of maintenance skill* (Research Report No. ADA1201672XSP). Alexandria, VA: Army Research Institute.

Hall, E. R., Ford, L. H., Whitten, T. C., & Plyant, L. R. (1983). *Knowledge retention among graduates of basic electricity and electronics schools* (HumRRO Tech. Rep 149). Orlando, FL: Training Analysis and Evaluation Group, Department of the Navy.

Hanges, P. J., Schneider, B., & Niles, K. (1990). Stability of performance: An interactionist perspective. *Journal of Applied Psychology, 75*, 658-667.

Hays, W. L. (1988). *Statistics*. New York: Holt Rinehart.

Holgrem, J. R., Hilligoss, R., Swezey, R. W., & Enkins, R. (1979). *Training effectiveness and retention of training extension course instruction in the combat arms* (Research Report 1208). Alexandria, VA: Army Research Institute.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin, 107*, 328-340.

Johnston, A. N. , Regian, J. W., & Shebilske, W. L. (1994). Observational learning and training of complex skill is laboratory and applied settings. *Proceedings of the 21st Conference of the Western European Association for Aviation Psychology*.

Jordan, J. A. (1991). *Modeling in dyadic protocols for training complex skills*. Unpublished master's thesis. Texas A&M University, College Station, TX.

Kirkpatrick, D. L. (1987). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 301-319). New York: McGraw-Hill.

Kyllonen, P. C., Christal, R. E., Woltz, D.J., Shute, V. J., Tirre, W. C., & Chaiken, S. (1990). *Cognitive abilities measurement (CAM) test battery: Version 4.0* [Computer program]. Brooks Air Force Base, TX: Armstrong Laboratory.

Logg, E. (1993). *Microsoft Arcade: Asteroids (Computer program)*. Redmond, WA: Microsoft Corporation.

Mane, A. M. & Donchin, E. (1989). The Space Fortress game. *Acta Psychologica, 71*, 17-22.

Mihal, W., & Barrett, G. V. (1976). Individual differences in perceptual information processing and their relation to automobile accident involvement. *Journal of Applied Psychology, 61*, 229-233.

Pintrich, P. R., Cross, D. R., Kozma, R. B., & McKeachie, W. J. (1986). Instructional psychology. *Annual Review of Psychology, 37*, 611-651.

Prislin, R., Jordan, J. A., Worchel, S., Tschan Semmer, F., & Shebilske, W. L. (1996). Effects of group discussion on acquisition of complex skills. *Human Factors, 38*, 404-416.

Purdy, B. J., & Lockhart, A. (1962). Retention and relearning of gross motor skills after long periods of no practice. *Research Quarterly, 33*, 265-272.

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space Fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta-Psychologica, 71*, 243-257.

Raven, J. C., Court, J. H., & Raven, J. (1993). *A manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67*, 605-610.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1*, 115-129.

Schmidt, F. L., & Hunter, J. E. (1978). Moderator research and the law of small numbers. *Personnel Psychology, 31*. 215-232.

Schmidt, R. A., & Björk, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts in training. *Psychological Science, 3*, 207-217.

Shebilske, W. L., Goettl, B., & Regian, J. W. (in press). Individual and group protocols for training complex skills in laboratory and applied settings. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*. Hillsdale, NJ: Erlbaum.

Shebilske, W. L., Jordan, J. A. & Arthur, W., Jr. (1993). Cognitive factors in automated instruction for individuals and groups (Abstract). *Proceedings of the 34th Annual Meeting of the Psychonomic Society, 1*, 24.

Shebilske, W. L., Jordan, J. A. Arthur, W., Jr., & Regian, J. W., (1993). Combining a multiple emphasis on components protocol with small group protocols for training complex skills. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society, 2*, 1216-1220.

Shebilske, W. L., & Regian, J. W., (1992). Video games, training, and investigating complex skills. *Proceedings of the 36th Annual Meeting of the Human Factors and Ergonomics Society, 2*, 1296-1300.

Shebilske, W. L., Regian, J. W., Arthur, W., Jr., & Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34*, 369-374.

Shields, J. L., Goldberg, S. L., & Dressel, J. D. (1979). *Retention of basic soldiering skills* (Research Report 1225). Alexandria, VA: U.S. Army Research Institute.

Singley, M., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard Univ. Press.

Snow, R. E. (1989). Cognitive-conative aptitude interactions in learning. In R. Kanfer & P. L. Ackerman (Eds.), *Ability, motivation, and methodology: The Minnesota symposium on learning and individual differences* (pp. 435-474). Hillsdale, NJ: Erlbaum.

Strong, M. H. (1992). *An assessment of the criterion-related validity and improvements in the utility of the Computer Administered Visual Selective Attention Test*. Unpublished masters thesis, Texas A&M University, College Station.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*, 399-441.

Theurer, D. (1993). *Microsoft Arcade: Tempest (Computer program)*. Redmond, WA: Microsoft Corporation.

Tubre, T. C. (1997). *The role of social factors in a dyadic training protocol: Implications for computer-based instruction* Unpublished master's thesis. Texas A&M University, College Station, TX.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

Vineberg, R. (1975). *A study of the retention skills and knowledge acquired in basic training* (HumRRO TR 75-10). Alexandria, VA: Human Resources Research Organization.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Woehr, D. J., & Miller, M. J. (1997). *The meaning and measurement of work ethic.* Paper presented at the annual Southeastern Industrial/Organizational Psychology Meeting, Atlanta, GA.

APPENDIX A - Correlations Among Predictors

Predictors	1	2	3	4	5	6	7	8	9	10	11	12	13
1. CA-VAT-I	1.00	.70	-.11	-.21	-.27	-.38	-.42	.12	-.43	-.45	.22	-.43	.11
2. CA-VAT-II	.70	1.00	-.03	-.31	-.35	-.28	-.30	.17	-.35	-.60	.37	-.61	.18
3. CATT	-.11	-.03	1.00	.05	.19	.07	.19	.08	.12	-.04	.03	-.04	.21
4. DKT-I	-.21	-.31	.05	1.00	.61	.05	.14	-.10	.18	.32	-.11	.29	-.16
5. DKT-II	-.27	-.35	.19	.61	1.00	.75	.22	-.02	.21	.31	-.16	.31	-.03
6. DKT-III	-.38	-.28	.07	.45	.75	1.00	.28	.02	.24	.29	-.09	.26	-.03
7. *g*	-.42	-.30	.19	.14	.22	.28	1.00	.06	.85	.33	.04	.24	.04
8. *g*-RT	.12	.17	.08	-.10	-.02	.02	.06	1.00	-.49	-.04	.44	-.20	.15
9. *g*-RTRW	-.43	-.35	.12	.18	.21	.24	.85	-.49	1.00	.31	-.20	.32	-.04
10. SWMT	-.45	-.60	-.04	.32	.31	.29	.33	-.04	.31	1.00	-.40	.93	-.05
11. SWMT-RT	.22	.37	.03	-.11	-.16	-.09	.04	.44	-.20	-.40	1.00	-.70	.05
12. SWMT-RTRW	-.43	-.61	-.04	.29	.31	.26	.24	-.20	.32	.91	-.70	1.00	-.09
13. GUM-A	.11	.18	.21	-.16	-.03	-.03	.04	.15	-.04	-.09	.05	-.09	1.00
14. GUM-C	.01	.09	.39	-.02	.00	.06	.12	.24	-.03	.02	.16	-.05	.49
15. GUM-E	-.04	-.11	.02	-.11	-.07	-.10	.25	.19	.12	.06	.02	.04	.17
16. GUM-O	-.07	-.03	.16	.02	-.06	.03	-.02	.03	-.03	.09	-.09	.10	.21
17. GUM-ES	-.08	.04	-.06	-.11	-.11	-.04	.09	.20	-.03	.03	.12	-.02	.43
18. NEO-A	.1	.03	.20	-.16	-.10	-.08	.11	.12	.03	.08	-.04	.08	.75
19. NEO-C	-.10	-.04	.37	.01	-.03	.06	.15	.10	.08	-.01	.12	-.05	.39
20. NEO-E	.05	-.08	.07	-.04	.02	-.00	.28	.09	.20	.05	.06	.02	.44
21. NEO-O	-.19	-.08	.18	.10	-.04	-.06	.06	-.18	.15	.17	-.22	.22	-.04
22. NEO-N	.13	.15	-.07	-.01	.11	.07	-.11	-.08	-.05	-.07	-.06	-.03	-.27
23. TR-COMM	-.07	.05	-.03	-.09	-.12	-.03	.07	-.03	-.08	.10	.07	.05	.07
24. NTR-COMM	-.05	-.13	.04	.00	-.07	-.17	.17	-.01	.16	.17	-.13	.19	.14
25. SPST	-.29	-.04	.11	.13	.07	.07	.19	.04	.15	.25	.07	.16	-.06
26. SPST-RT	.40	.38	.15	-.22	-.29	-.33	-.18	.11	-.22	-.47	.35	-.50	.14
27. SPST-RTRW	-.39	-.18	.03	.19	.17	.19	.23	-.01	.20	.38	-.08	.33	-.10
28. AF-ACCD	-.13	-.03	-.03	.05	-.02	-.06	.08	.10	.02	-.04	.12	-.08	.18
29. IAS	.13	.17	.08	-.11	.05	.10	-.21	-.05	-.16	-.09	.01	-.07	-.11
30. MCSD	.06	.07	.05	-.01	-.07	.07	.09	.06	.04	-.05	.05	-.06	.45
31. SMS	-.15	-.08	-.01	-.06	-.08	-.14	.08	.09	.02	-.01	-.09	.02	.04
32. USRT-I	.27	.25	-.12	-.04	.02	-.05	-.00	.03	-.02	-.10	.09	-.12	.05
33. USRT-II	.16	.19	-.05	.01	.04	-.04	.05	-.02	.05	-.08	.01	-.06	.07
34. USRT-III	.14	.15	-.20	-.05	.05	-.08	.00	.06	-.04	-.09	-.04	-.05	-.05
35. USRT-IV	.07	.17	-.21	-.04	-.01	-.17	-.08	-.07	-.04	-.11	-.10	-.04	-.17
36. USRT-V	.11	.06	-.22	.02	.05	-.19	-.05	-.04	-.02	.00	-.12	.04	-.10
37. USRT-VI	.05	.05	-.28	.04	.09	-.08	-.01	.05	-.03	-.02	-.09	.02	-.12
38. USRT-VII	.03	.04	-.21	.06	.12	-.05	-.08	-.01	-.06	-.05	-.17	.02	-.18

APPENDIX A - Correlations Among Predictors - Continued

Predictors	14	15	16	17	18	19	20	21	22	23	24	25	26
1. CA-VAT-I	.01	-.04	-.07	-.08	.01	-.10	.05	-.19	.13	-.07	-.05	-.29	.40
2. CA-VAT-II	.09	-.11	-.03	.04	.03	-.04	-.08	-.08	.15	.05	-.13	-.04	.38
3. CATT	.39	.02	.16	-.06	.20	.37	.06	18	-.07	-.03	.04	.11	.15
4. DKT-I	-.02	-.11	.02	-.11	-.16	01	.04	.10	-.01	-.09	.00	.13	-.22
5. DKT-II	.00	-.07	-.06	-.11	-.10	-.03	02	.04	.11	.12	-.07	.07	-.29
6. DKT-III	.06	-.10	.03	-.04	-.08	.06	-.00	.06	07	-.03	-.17	.07	-.33
7. *g*	.12	.25	-.02	.09	.11	.15	.28	.06	-.11	.07	.17	.19	-.18
8. g-RT	.24	.19	.03	.20	.12	.10	.09	-.18	-.08	.03	-.01	.04	.11
9. g-RTRW	-.03	.12	-.03	-.03	.03	.08	.20	.15	-.05	.08	.16	.15	-.22
10. SWMT	.02	.06	.09	.03	.08	-.01	.05	.17	-.07	.10	.17	.25	-.47
11. SWMT-RT	.16	.02	-.09	.12	-.04	.12	06	.22	-.06	07	-.13	.07	.35
12. SWMT-RTRW	-.05	.04	.10	-.03	.08	-.05	.02	22	-.03	.05	.19	.16	-.50
13. GUM-A	.49	.17	.21	.43	.75	.39	.44	-.04	-.27	.07	.14	-.06	.14
14. GUM-C	1.00	.00	.12	.20	.45	.84	.18	-.15	-.22	.12	.10	.02	.04
15. GUM-E	.00	1.00	.30	.26	.13	-.03	.67	.04	-.26	.01	.12	-.04	.01
16. GUM-O	.12	.30	1.00	.06	.01	.12	.14	.59	.04	.02	.01	-.06	-.10
17. GUM-ES	.20	.26	.06	1.00	.35	.05	15	-.03	-.60	.16	.18	.19	-.08
18. NEO-A	.45	.13	.01	.35	1.00	.38	.34	-.09	-.38	.06	.28	-.01	.10
19. NEO-C	.84	-.03	.12	.06	.38	1.00	.18	-.10	-.27	.06	.10	-.05	.10
20. NEO-E	.18	.67	.14	.12	.34	.18	1.00	-.07	-.28	-.06	.15	-.03	.12
21. NEO-O	-.15	.04	.59	-.03	-.09	-.10	-.07	1.00	.07	.04	.05	.02	-.05
22. NEO-N	-.22	-.26	-.04	-.60	-.38	-.27	-.28	.07	1.00	-.05	-.17	-.09	.04
23. TR-COMM	.12	.01	.02	.16	.06		-.06	.04	.05	1.00	.19	.15	-.16
24. NTR-COMM	.10	.12	.01	.18	.28	.10	.15	.05	-.17	.19	1.00	.07	-.02
25. SPST	.02	-.04	-.06	.19	-.01	-.05	-.03	.02	-.09	.15	.07	1.00	-.29
26. SPST-RT	.04	.01	-.10	-.08	.10	.10	.12	-.05	.04	-.16	-.02	-.28	1.00
27. SPST-RTRW	.00	-.04	-.02	.19	-.04	-.08	-.07	.03	-.09	.18	.06	.93	-.62
28. AF-ACCD	.06	.05	-.04	.31	.22	.02	.06	.04	-.15	.01	.01	.30	.04
29. IAS	.02	-.70	-.33	-.34	-.06	-.02	-.41	-.18	.48	.03	-.15	.08	.09
30. MCSD	.33	.03	.11	.33	.40	.39	.14	-.14	-.34	.02	.07	.02	-.04
31. SMS	-.10	.41	.16	.10	-.05	-.12	.34	.13	-.03	-.03	-.11	-.17	.03
32. USRT-I	-.03	.00	-.10	-.04	.02	.04	11	-.18	.10	-.07	.13	-.08	.08
33. USRT-II	.00	-.04	-.04	-.01	.036	.08	.09	-.07	.14	-.01	.18	.02	.06
34. USRT-III	-.14	.12	-.04	.05	-.04	-.10	05	-.09	.17	-.09	.12	-.04	.05
35. USRT-IV	-.28	.09	.03	.00	-.18	-.18	-.04	.09	.12	-.02	.09	.01	.03
36. USRT-V	-.18	.11	.00	-.06	-.07	-.11	06	.01	.13	-.09	.08	-.01	.00
37. USRT-VI	-.17	.11	.02	.00	-.11	-.11	.04	-.01	.07	-.03	.03	.06	.00
38. USRT-VII	-.28	.15	.04	.01	-.15	-.23	.03	.12	.04	.00	.06	.03	.01

APPENDIX A - Correlations Among Predictors - Continued

Predictors	27	28	29	30	31	32	33	34	35	36	37	38
1. CA-VAT-I	-.39	-.13	.13	.06	-.15	.27	.16	.14	.07	.11	.05	.11
2. CA-VAT-II	-.18	-.03	.17	.07	-.08	.25	.19	.15	.17	.06	.05	.18
3. CATT	.03	-.03	.08	.05	-.01	-.12	-.05	-.20	-.21	-.22	-.28	.21
4. DKT-I	.19	.05	-.11	-.01	-.06	-.04	.01	-.05	-.04	.02	.04	-.16
5. DKT-II	.17	-.02	.05	-.09	-.08	.02	.04	.05	-.01	.05	.09	-.03
6. DKT-III	.19	-.06	.10	.07	-.14	-.05	-.04	-.08	-.17	-.19	-.08	-.03
7. *g*	.23	.08	-.21	.09	.08	.00	.04	.00	-.08	-.05	-.01	.04
8. *g*-RT	-.01	.10	-.05	.06	.09	.03	-.02	.06	-.07	-.04	.05	.15
9. *g*-RTRW	.20	.02	-.16	.04	.02	-.02	.05	-.04	-.04	-.02	-.03	-.04
10. SWMT	.38	-.04	-.09	-.05	-.01	-.10	-.08	-.09	-.11	.00	-.02	-.09
11. SWMT-RT	-.08	.12	.01	.05	-.09	.09	.01	-.04	-.10	-.12	-.09	.05
12. SWMT-RTRW	.33	-.08	-.07	-.06	.02	-.12	-.06	-.05	-.04	.04	.02	-.09
13. GUM-A	-.10	.18	-.11	.45	.04	.05	.07	-.05	-.17	-.10	-.12	1.00
14. GUM-C	.00	.03	.02	.33	-.10	-.03	.00	-.14	-.28	-.18	-.17	.49
15. GUM-E	-.04	.05	-.69	.03	.41	.00	-.04	.12	.09	.11	.11	.17
16. GUM-O	-.01	-.04	-.33	.11	.16	-.10	-.04	-.04	.03	.00	.02	.21
17. GUM-ES	.19	.31	-.38	.33	.10	-.04	-.01	.05	.00	-.06	-.01	.43
18. NEO-A	-.04	.22	-.06	.39	-.05	.02	.03	-.04	-.18	-.07	-.11	.75
19. NEO-C	-.08	.02	-.02	.39	-.12	.04	.08	-.10	-.18	-.11	-.11	.39
20. NEO-E	-.07	.06	-.41	.14	.34	.11	.09	.05	-.04	.06	.04	.44
21. NEO-O	.03	.04	-.18	-.14	.13	-.18	-.07	-.09	.09	.01	-.01	-.04
22. NEO-N	-.09	-.15	.48	-.34	-.03	.10	.14	.17	.12	.13	.07	-.27
23. TR-COMM	.18	.00	.03	.02	-.03	-.07	-.01	-.09	-.01	-.09	-.03	.07
24. NTR-COMM	.06	.01	-.15	.07	-.11	.13	.18	.12	.09	.08	.03	.14
25. SPST	.93	.30	.08	.02	-.17	-.08	.02	-.04	.01	-.01	.06	-.06
26. SPST-RT	-.62	.04	.09	-.04	.03	.08	.06	.05	.03	.00	.00	.14
27. SPST-RTRW	1.00	.23	.03	.04	-.15	-.09	-.01	-.06	.00	-.01	.05	-.10
28. AF-ACCD	.23	1.00	-.17	-.02	.17	.03	.00	-.05	-.01	.02	.10	.18
29. IAS	.03	-.17	1.00	-.12	-.25	-.05	.08	-.03	-.12	-.06	-.09	-.11
30. MCSD	.04	-.02	-.12	1.00	-.25	.13	.06	.05	-.12	-.10	-.04	-.23
31. SMS	-.15	.17	-.25	-.25	1.00	-.07	-.05	.18	.15	.20	.16	.29
32. USRT-I	-.09	.03	-.05	.13	-.07	1.00	.66	.59	.48	.52	.44	.33
33. USRT-II	-.01	.00	.08	.06	-.05	.66	1.00	.75	.61	.63	.57	.47
34. USRT-III	-.06	-.05	-.03	.05	.18	.59	.75	1.00	.78	.75	.70	.66
35. USRT-IV	.00	-.01	-.12	-.12	-.15	.48	.61	.78	1.00	.77	.69	.78
36. USRT-V	-.01	.02	-.06	-.10	.20	.52	.63	.75	.77	1.00	.80	.70
37. USRT-VI	.05	.10	-.09	-.04	.16	.44	.57	.70	.69	.80	1.00	.73
38. USRT-VII	.02	.20	-.17	-.23	.29	.33	.47	.66	.78	.70	.73	1.00

Legend

1.	CA-VAT-I	Computer-Administered Visual Attention Test - first administration
2.	CA-VAT-II	Computer-Administered Visual Attention Test - second administration
3.	CATT	Computer Attitude Scale
4.	DKT-I	Declarative Knowledge Test - first administration
5.	DKT-II	Declarative Knowledge Test - second administration
6.	DKT-III	Declarative Knowledge Test - third administration
7.	*g*	Figure Matrices Test
8.	*g*-RT	Figure Matrices Test - reaction time
9.	*g*-RTRW	Figure Matrices Test - reaction time and right/wrong algorithm
10.	SWMT	Spatial Working Memory Test
11.	SWMT-RT	Spatial Working Memory Test - reaction time
12.	SWMT-RTRW	Spatial Working Memory Test - reaction time and right/wrong algorithm
13.	GUM-A	Goldberg 100 Unipolar Markers - agreeableness
14.	GUM-C	Goldberg 100 Unipolar Markers - conscientiousness
15.	GUM-E	Goldberg 100 Unipolar Markers - extraversion
16.	GUM-O	Goldberg 100 Unipolar Markers - openness
17.	GUM-ES	Goldberg 100 Unipolar Markers - emotional stability
18.	NEO-A	NEO Five-Factor Inventory - agreeableness
19.	NEO-C	NEO Five-Factor Inventory - conscientiousness
20.	NEO-E	NEO Five-Factor Inventory - extraversion
21.	NEO-O	NEO Five-Factor Inventory - openness
22.	NEO-N	NEO Five-Factor Inventory - neuroticism
23.	TR-COMM	Task-Related Communication
24.	NTR-COMM	Non-Task-Related Communication
25.	SPST	Spatial Processing Speed Test
26.	SPST-RT	Spatial Processing Speed Test - reaction time
27.	SPST-RTRW	Spatial Processing Speed Test - reaction time and right/wrong algorithm
28.	AF-ACCD	At-Fault Accidents
29.	IAS	Interaction Anxiousness Scale
30.	MCSD	Marlowe/Crowne Social Desirability Scale
31.	SMS	Self-Monitoring Scale
32.	USRT-I	Unprepared Simple Reaction Time Test - first administration
33.	USRT-II	Unprepared Simple Reaction Time Test - second administration
34.	USRT-III	Unprepared Simple Reaction Time Test - third administration
35.	USRT-IV	Unprepared Simple Reaction Time Test - fourth administration
36.	USRT-V	Unprepared Simple Reaction Time Test - fifth administration
37.	USRT-VI	Unprepared Simple Reaction Time Test - sixth administration
38.	USRT-VII	Unprepared Simple Reaction Time Test - seventh administration

APPENDIX B - Correlations Among Criteria

Criteria	1	2	3	4	5	6	7	8	9	10	11	12
1. Aiming Task	1.00	.39	.37	.45	.44	.45	.39	.09	.26	.16	.31	.32
SPACE FORTRESS												
2. Baseline	.39	1.00	.65	.60	.60	.54	.53	.36	.30	.29	.36	.24
3. Session 4	.37	.65	1.00	.79	.74	.69	.69	.18	.59	.44	.44	.45
4. Session 9	.45	.60	.79	1.00	.85	.81	.86	.09	.46	.60	.58	.54
5. Session 10	.44	.60	.74	.85	1.00	.85	.83	.12	.41	.56	.64	.51
6. Session 14	.45	.54	.69	.81	.85	1.00	.86	.07	.47	.46	.60	.61
7. Session 19	.39	.53	.69	.86	.83	.86	1.00	.09	.45	.54	.62	.66
STRATEGY												
8. Baseline	.09	.36	.18	.09	.12	.07	.09	1.00	.12	.03	.13	-.05
9. Session 4	.26	.30	.59	.46	.41	.47	.45	.12	1.00	.37	.24	.39
10. Session 9	.16	.29	.44	.60	.56	.46	.54	.03	.37	1.00	.65	.37
11. Session 10	.31	.36	.44	.58	.64	.60	.62	.13	.24	.65	1.00	.48
12. Session 14	.32	.24	.45	.54	.51	.61	.66	-.05	.39	.37	.48	1.00
13. Session 19	.12	.27	.39	.55	.53	.60	.63	-.09	.45	.68	.64	.50
ASTEROIDS												
14. First Session	.23	.19	.22	.24	.17	.18	.20	-.03	.07	-.04	.04	.02
15. Second Session	.27	.25	.21	.13	.13	.12	.08	.14	.12	.09	.16	.20
16. Third Session	.29	.35	.34	.35	.40	.29	.34	.03	.13	.21	.25	.30
17. Fourth Session	.24	.25	.26	.35	.46	.37	.41	.12	.12	.28	.41	.21
TEMPEST												
18. First Session	.01	.10	.11	.17	.04	.11	.11	.08	.07	.03	-.07	-.07
19. Second Session	-.02	.07	-.06	-.08	-.09	-.12	-.12	.12	-.04	-.25	-.03	-.19
20. Third Session	.08	.05	.07	.04	-.00	-.03	-.02	.14	.16	-.10	-.09	-.03
21. Fourth Session	.13	.10	.12	.12	.13	.15	.17	.07	.19	.02	-.09	.06
SF KEYBOARD												
22. Session 12	.36	.41	.55	.62	.64	.62	.60	.02	.44	.43	.47	.51
23. Session 18	.30	.28	.40	.50	.51	.47	.51	-.14	.25	.28	.32	.40

APPENDIX B - Correlations Among Criteria - Continued

Criteria	13	14	15	16	17	18	19	20	21	22	23
1. Aiming Task	.12	.23	.27	.29	.24	.01	-.02	.08	.13	.36	.30
SPACE FORTRESS											
2. Baseline	.27	.19	.25	.35	.25	.10	.07	.05	.10	.41	.28
3. Session 4	.39	.22	.21	.34	.26	.11	-.06	.07	.12	.55	.40
4. Session 9	.55	.24	.13	.34	.35	.17	-.08	.04	.14	.62	.50
5. Session 10	.53	.17	.13	.40	.46	.04	-.09	-.00	.13	.64	.51
6. Session 14	.60	.18	.12	.29	.37	.11	-.12	-.03	.15	.62	.47
7. Session 19	.63	.20	.08	.34	.41	.11	-.12	-.02	.17	.60	.51
STRATEGY											
8. Baseline	-.09	-.03	.14	.03	.12	.08	.12	.14	.07	.02	-.14
9. Session 4	.45	.07	.12	.13	.12	.07	-.04	.16	.19	.44	.25
10. Session 9	.68	-.04	.09	.21	.28	.03	-.25	-.10	.02	.43	.28
11. Session 10	.64	.04	.16	.25	.41	-.07	-.03	-.09	-.09	.47	.32
12. Session 14	.50	.02	.20	.30	.21	-.07	-.19	-.03	.06	.51	.40
13. Session 19	1.00	-.03	.02	.18	.33	.04	-.24	-.05	-.05	.40	.32
ASTEROIDS											
14. First Session	-.03	1.00	.17	.29	.24	.60	.15	.22	.10	-.10	-.02
15. Second Session	.02	.17	1.00	.39	.21	-.02	-.12	-.11	.02	.02	.11
16. Third Session	.18	.29	.30	1.00	.49	.08	.03	.11	.23	.23	.26
17. Fourth Session	.33	.24	.21	.49	1.00	.05	.04	.01	.21	.21	.20
TEMPEST											
18. First Session	.04	.60	-.02	.08	.05	1.00	.16	.28	.06	-.16	-.24
19. Second Session	-.24	.15	-.12	.03	.04	.16	1.00	.39	.02	-.12	-.11
20. Third Session	-.05	.22	-.11	.11	.01	.28	.39	1.00	.11	.08	.05
21. Fourth Session	-.05	.10	.02	.23	.21	.06	.02	.11	1.00	.16	.12
SF KEYBOARD											
22. Session 12	.40	-.10	.02	.23	.21	-.16	-.12	.08	.16	1.00	.73
23. Session 18	.32	-.02	.11	.26	.20	-.24	-.11	.05	.12	.73	1.00

APPENDIX C - Correlations Among Predictors and Criteria

	Predictors												
Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Aiming Task	-.07	-.12	.15	.10	.05	.10	.11	-.06	.13	.23	-.29	29	15
SPACE FORTRESS													
2. Baseline	-.25	-.30	.26	.33	.30	.28	.21	-.02	.19	.19	-.12	.20	-.07
3. Session 4	-.30	-.32	.15	.36	.42	.37	.34	-.09	.35	.17	-.13	.19	-.03
4. Session 9	-.31	-.36	.12	.32	.37	.38	.24	-.09	.26	.13	-.15	.16	.01
5. Session 10	-.28	-.32	.03	.37	.38	.41	.21	-.12	.25	.21	-.21	.24	.08
6. Session 14	-.25	-.30	.12	.28	.34	.35	.16	-.08	.18	.13	-.24	.20	.02
7. Session 19	-.31	-.29	.01	.29	.35	.44	.17	-.07	.18	.15	-.16	.18	.01
STRATEGY													
8. Baseline	-.24	-.17	.18	-.04	.02	.07	.24	-.01	.22	.15	-.17	.18	.11
9. Session 4	-.24	-.26	.21	.19	.23	.18	.24	-.02	.22	.19	-.05	.17	.01
10. Session 9	-.29	-.18	-.02	.21	.28	.30	.13	-.18	.21	.11	-.07	.11	-.01
11. Session 10	-.34	-.19	.03	.21	.21	.32	22	-.08	.24	.13	-.11	.14	.09
12. Session 14	-.34	-.27	.01	.22	.25	.35	.12	.00	.11	.21	-.12	.21	-.18
13. Session 19	-.28	-.20	.09	.15	.25	.30	.12	-.16	.19	.00	-.12	.05	-.05
ASTERIODS													
14. First Session	.10	-.07	.09	.05	.01	.00	-.12	-.09	-.06	-.02	-.06	.01	.28
15. Second Session	-.01	-.09	.16	.13	.05	-.05	-.01	-.02	.01	.10	-.01	.09	.12
16. Third Session	-.12	.01	.13	.24	.03	.10	.05	-.12	.10	-.01	-.01	.12	.12
17. Fourth Session	-.13	-.14	-.06	.07	.04	.19	.06	-.15	.13	.16	-.25	-.06	.27
TEMPEST													
18. First Session	.16	-.02	-.02	-.26	-.14	-.13	-.15	-.07	-.09	-.22	-.20	-.10	.10
19. Second Session	-.05	.02	-.05	.02	.00	.08	.07	-.08	.10	-.15	-.04	-.10	.13
20. Third Session	-.05	-.01	.10	-.09	-.08	-.07	-.02	-.28	.13	-.05	-.20	.04	-.06
21. Fourth Session	.00	.03	.00	.03	-.05	.03	.07	-.09	.11	-.03	-.10	.01	-.13
SF KEYBOARD													
22. Session 12	-.30	-.29	-.04	.31	.25	.27	.31	-.13	.34	.22	-.10	.22	-.09
23. Session 18	-.27	-.21	-.04	.36	.36	.37	.26	-.03	.24	.22	-.08	.20	-.03

APPENDIX C - Correlations Among Predictors and Criteria - Continued

Criteria	Predictors 14	15	16	17	18	19	20	21	22	23	24	25	26
1. Aiming Task	.18	-.14	.10	.16	.19	.14	-.02	.13	-.20	-.04	.09	.07	-.25
SPACE FORTRESS													
2. Baseline	.03	-.14	.01	-.09	-.07	.06	-.10	.11	.04	-.12	.03	.11	-.26
3. Session 4	-.02	-.05	.00	-.06	.05	.02	.05	.06	-.10	-.23	.03	.04	-.29
4. Session 9	.01	-.19	.02	.02	-.01	.12	-.08	.08	-.11	-.20	.06	.06	-.28
5. Session 10	.02	-.13	-.03	.01	.10	.01	-.01	.12	-.07	-.24	.07	.05	-.24
6. Session 14	.00	-.13	-.01	.00	.05	.02	-.05	.11	-.03	-.23	.07	.00	-.29
7. Session 19	.02	-.11	-.03	.01	.03	.06	-.02	.11	-.04	-.03	-.01	.00	-.31
STRATEGY													
8. Baseline	.12	-.01	.16	.03	.13	.07	.01	.13	.00	.13	.08	.05	-.14
9. Session 4	.07	-.05	-.09	-.05	.14	.20	.07	.14	-.05	-.11	.00	.08	-.13
10. Session 9	.03	-.32	.15	-.09	-.05	.16	-.13	.08	.03	-.09	.13	.11	-.15
11. Session 10	-.05	-.20	-.03	.10	.08	-.01	-.04	.10	-.08	-.01	17	.14	-.14
12. Session 14	.05	-.25	-.07	.00	-.09	.02	-.23	.10	.00	-.01	-.06	.03	-.29
13. Session 19	-.05	-.17	-.19	-.01	-.06	.02	-.06	.01	-.03	.01	.14	.04	-.10
ASTERIODS													
14. First Session	.16	.05	.10	.19	.23	.09	.05	-.02	-.19	.03	-.09	-.08	-.08
15. Second Session	.28	-.07	-.04	-.02	.17	.23	-.03	-.08	-.07	.01	.02	.06	-.09
16. Third Session	.08	-.10	-.04	.11	.09	.04	-.01	.12	-.07	-.04	-.03	.15	.02
17. Fourth Session	.01	.04	.12	.04	.27	.02	.16	.17	.04	-.08	.08	-.12	-.10
TEMPEST													
18. First Session	-.03	-.06	-.10	.11	.11	-.04	.14	-.09	-.05	-.01	.02	-.17	-.04
19. Second Session	-.04	.06	.02	.00	.11	-.02	.17	-.05	.01	-.08	-.06	.09	.02
20. Third Session	-.14	-.06	-.05	.00	-.01	-.07	-.06	.09	.03	-.02	-.13	-.01	.00
21. Fourth Session	.01	.06	.03	-.07	-.02	.13	.01	.24	-.02	-.09	-.02	.00	-.08
SF KEYBOARD													
22. Session 12	-.06	-.26	-.12	.02	-.04	.06	-.09	.19	.03	-.05	.16	.15	-.31
23. Session 18	.02	-.14	-.15	.04	-.04	.08	-.02	-.03	.05	.05	.05	.21	-.27

APPENDIX C - Correlations Among Predictors and Criteria - Continued

Criteria	27	28	29	30	31	32	33	34	35	36	37	38
	Predictors											
1. Aiming Task	.16	-.09	-.11	.28	-.17	-.15	-.20	-.21	-.24	-.25	-.22	-.24
SPACE FORTRESS												
2. Baseline	.19	.00	.07	.13	-.03	-.29	-.25	-.29	-.31	-.29	-.30	-.31
3. Session 4	.15	.02	-.07	.17	.01	-.10	-.07	-.07	-.09	-.06	.00	-.09
4. Session 9	.16	.04	.03	.18	-.11	-.21	-.15	-.15	-.17	-.21	-.09	-.20
5. Session 10	.13	.06	.02	.11	-.08	-.20	-.25	-.25	-.27	-.23	-.20	-.25
6. Session 14	.11	.03	.06	.13	-.08	-.19	-.25	-.21	-.25	-.26	-.23	-.22
7. Session 19	.12	-.03	.06	.14	-.01	-.30	-.26	-.18	-.22	-.21	-.14	-.23
STRATEGY												
8. Baseline	.09	.05	.07	.05	.13	-.22	-.18	-.21	-.28	-.16	-.15	-.13
9. Session 4	.11	.00	.02	.09	.03	-.13	.04	-.04	-.05	.04	.01	-.03
10. Session 9	.15	.09	.18	-.03	-.13	-.11	-.01	-.06	.00	-.06	.01	-.03
11. Session 10	.17	.11	.12	.01	.03	-.16	-.13	-.05	-.08	-.16	-.12	-.02
12. Session 14	.14	-.04	.18	-.07	.08	-.13	-.13	-.02	-.05	-.05	-.05	-.09
13. Session 19	.07	.02	.16	-.05	-.08	-.20	-.12	-.13	-.08	-.20	-.18	-.08
ASTERIODS												
14. First Session	-.03	-.02	-.16	.27	-.08	-.08	-.05	-.08	-.11	-.13	-.24	-.25
15. Second Session	.09	-.04	.00	.16	-.05	-.14	-.11	.00	-.15	-.12	-.14	-.14
16. Third Session	.12	.06	.01	.17	-.07	-.10	-.04	-.09	-.08	-.12	-.11	-.26
17. Fourth Session	-.06	-.17	.04	.10	.01	-.05	-.05	-.04	-.10	-.11	-.15	-.19
TEMPEST												
18. First Session	-.13	.03	.03	.14	.04	-.13	-.07	.00	-.04	-.03	.00	-.06
19. Second Session	.07	.14	-.01	.15	.04	.05	.10	.01	-.01	-.01	.06	.04
20. Third Session	-.01	.04	-.04	.03	.02	-.05	-.08	-.08	.02	.02	-.07	-.04
21. Fourth Session	.03	.07	-.03	.09	.01	-.06	-.07	-.07	-.11	-.09	-.06	.07
SF KEYBOARD												
22. Session 12	.24	.07	.04	.14	-.07	.02	-.05	-.03	-.07	-.04	-.03	-.06
23. Session 18	.27	.06	.01	.17	-.24	.05	-.06	-.04	-.16	-.13	-.12	-.22

APPENDIX D - ADDITIONAL MEASURES PRESENTED IN APPENDICES A-C BUT NOT USED IN THIS REPORT

Computer Attitude Scale (CATT). This is a 20-item instrument that measures the respondent's attitudes toward computers. This measure was computer-administered. Given the administration and responding format of most of the measures used in the present study, the CATT (Dambrot, Watkins-Malek, Silling, Marshall, & Garver, 1985) was administered to permit the assessment of the effect of computer attitudes on the other computer-administered measures. Internal consistencies of .84, .79, and .73 have been reported for the CATT (Arthur & Olson, 1991; Dambrot et al., 1985).

Goldberg's 100 Unipolar Markers (Goldberg, 1992). This is a measure of the Big Five personality factors. Using a 9-point scale (1 = extremely inaccurate, to 9 = extremely accurate), subjects rate a list of 100 common human traits in terms of how accurately they describe themselves. This measure was computer-administered. Each factor is measured by 20 items and a subject's score on each factor is their mean rating on the 20 items. The average time to completion is approximately 15 minutes. For self assessments, internal consistencies of .90 (Extraversion), .88 (Agreeableness), .90 (Conscientiousness), .84 (Emotional Stability), and .85 (Openness) have been reported for the five factors (Goldberg, 1992). Goldberg (1992) also reports convergent validities of .69, .56, .67, -.69, and .46 between the NEO-FFI and the 100 Unipolar Markers' assessment of the Big Five.

NEO Five-Factor Inventory (NEO-FFI) (Form S; Costa & McCrae, 1985). This is a 60-item measure of the Big Five, with 12 items per personality dimension. This measure was computer-administered. Items are responded to on a five-point Likert scale (strongly disagree to strongly agree) and the average time to completion is approximately 15 minutes. Internal consistencies for the NEO-FFI scales as specified in the test manual are .79 (Extraversion), .74 (Agreeableness), .84 (Conscientiousness), .89 (Neuroticism), and .76 (Openness) (Costa & McCrae, 1991). Six-year test-retest reliabilities of .82, .83, and .83 for Extraversion, Neuroticism, and Openness, along with three-year retest reliabilities of .63 and .79 for Agreeableness and Conscientiousness are also reported in the test manual (Costa & McCrae, 1991).

Task-Related Communication. The structure of the AIM-Dyad protocol requires trainees to perform each half of a task alternately with a partner who performs the other half. The goal is for each trainee to learn both parts by hands-on practice on alternate trials and to learn the connection between parts by modeling the actions and reactions of their partner. To facilitate this, dyad members have always been strongly encouraged to communicate about the task. For example, the mine/missile manager is advised to tell the pilot/gunner whether the mines are friend or foe and when the ship's missiles would be effective against the mines. However, to date, no empirical data has been formally collected to assess how much communication actually takes place between them.

The present study sought to address this limitation by unobtrusively recording, and then scoring the amount of verbal communication that subjects' engaged in. Although dyadic subjects were, again, strongly encouraged to communicate, subjects were not informed that their communications would be recorded. Recording communications was accomplished by placing the microphones behind a barrier under the top shelf of the testing station. The tape recorders to which the microphones were connected, were hidden and located at the rear of the lab.

The unit of analysis that was coded and scored was operationally defined as "sentences or meaningful phrases" or utterances (Carrier & Sales, 1987). Four types of units were coded - these were specifically (a) task-specific/on-task; (b) task-elicited; (c) procedural; and (d) non-task/off-task communications. Task-specific/on-task communication was defined as communication that would directly aid the subject in performing the task. Examples of task-specific/on-task communication include "friend mine", "foe mine", and "we need missiles". Task-elicited communication was defined as communication relating to the task, but not necessarily one that would aid the subject in performing the task. Examples of task-elicited communication include "good speed", "are you ready?", and "nice wrapping". Procedural communication was defined as communication referring to the process of participating in the experiment. Examples of procedural communication include "when do you think we'll get paid", "the computer locked-up", "let's get the proctor", and "we're supposed to switch between games". Finally, non-task/off-task communication was defined as communication that was personal in nature, and completely disengaged from the task itself. Examples of non-task/off-task communication include "how was your weekend", "zippy, the wonder twit", and "finals are in two weeks".

Video Game Experience Questionnaire - Addendum A & B. Two short questionnaires (5 and 3 items) were developed to collect data on the extent to which subjects had played the two transfer tasks (i.e., Asteroids and Tempest) before signing up to participate in the study (Addendum A), and also during the eight week non-practice interval (Addendum B). Addendum A was attached to the Video Game Experience questionnaire (which was administered during the Screening). Both measures also asked subjects to rate their ability levels on the transfer tasks/games on a five-point Likert scale (1=novice, 3=average, 5=expert).

Driving Behavior Questionnaire. Subjects completed a driving behavior questionnaire (Arthur, 1991; Arthur & Doverspike, 1992). The questionnaire was computer-administered. In completing the questionnaire, subjects reported the total number of accidents they had been involved in as one of the drivers, the number for which they were at fault, and the total number of years they had been driving legally. An accident was defined as any driving or traffic accident in which the subject was involved as one of the drivers, and in which a person had suffered physical injury (including fatalities) and/or there was $150 or more damage to property. An at-fault accident was one in which the police had determined that the subject was at fault. Subjects were also asked to report the total number of moving violation tickets they had received. Arthur (1991) reports a total-accidents test-retest reliability of .96. Arthur and Graziano

(1994) obtained 2-3 day test-retest reliabilities of .98, .96, and .97 for total, at-fault, and not-at fault accidents respectively.

Confidence/Alertness Questionnaire (CAQ). Given the length of the Space Fortress training sessions used in the present study, the CAQ, which was designed to measure the subjects' fatigue, motivation, and confidence in their performance on the Space Fortress task, was included in the test battery. The CAQ is computer-administered and consists of 12 self-report items that are answered on a seven-point Likert scale. The two fatigue items were based on Folkard, Monk, and Lobban's (1978) work on shift work. The confidence items measured confidence in performance (present and future) relative to the other subjects, and the subject's own personal improvement.

Interaction Anxiousness Scale (IAS). The IAS (Leary, 1983) was designed to measure the tendency to experience subjective social anxiety independently of accompanying behaviors. The IAS, which was computer-administered, consists of 15 self-report items that are answered on a five-point Likert scale (1="not at all characteristic of me; 5="extremely characteristic of me"). Scale scores range from 15 (low interaction anxiety) to 75 (high interaction anxiety). Leary and Kowalski (1987) report a mean of 38.90 (*SD*=9.70) for a sample of 1140 college students. A coefficient alpha of .88 and an eight-week test-retest reliability of .80 have been reported for the IAS (Leary, 1983). In terms of its construct validity, IAS scores have been found to correlate highly (*r* values $>$ .60) with other measures of social anxiousness and shyness (Jones, Briggs, & Smith, 1986; Leary & Kowalski, 1987). Further, scores on the IAS correlate well with self-reported anxiety in real interactions. Compared to low scorers, high scorers report more anxiety and feel more inhibited during conversation (Leary, 1983, 1986). Finally, IAS scores correlate positively with social avoidance and inhibition (Leary, Atherton, Hill, & Hur, 1986).

Marlowe/Crowne Social Desirability Scale (MCSD). The MCSD (Crowne & Marlowe, 1964) measures a general concept of avoidance of disapproval. Given the number and nature of the self-report measures (e.g., the personality measures) used in the present study, the MCSD was included in the test battery to permit an assessment of the extent to which subjects were responding to these measures in a socially desirable manner, if such assessment was deemed to be warranted. The MCSD, which was computer-administered, consists of 33 items that are answered by responding "True" or "False" to 18 items keyed in the true direction and 15 items keyed in the false direction. Scores range from 0 (low need for approval) to 33 (high need for approval). Crowne and Marlowe (1964) report a mean of 15.5 (*SD*=4.4) for a sample of 300 college students, while Paulhus (1984) reports means of 13.3 (*SD*=4.3) for anonymous conditions, and 15.5 (*SD*=4.6) in public disclosure conditions. Reported coefficient alphas range from .73 to .88 with a one month test-retest reliability of .88 (Crowne & Marlowe, 1964).

The Self-Monitoring Scale (SMS). The SMS (Snyder, 1987) is an 18-item instrument that measures general differences in how people monitor their expressive behavior and self-presentation. Those found to be high self-

monitors tend to adopt interpersonal orientations in response to the situation presented. Those found to be low self-monitors, on the other hand, adopt interpersonal orientations in response to their own feelings, attitudes, and behaviors. This measure was included in the battery to determine the effects of self-monitoring on communication and interaction patterns within dyads. The SMS was computer-administered. The items are answered by responding "True" or "False" to the 8 items keyed in the true direction and 10 items keyed in the false direction. Scores range from 0 (low self-monitor) to 18 (high self-monitor). An internal consistency (coefficient alpha) of .70 has been reported for the SMS (Snyder, 1987).

Unprepared Simple Reaction Time Test (USRT). The USRT (Wilkinson & Houghton, 1982) was developed as a predictor of fatigue due to sleep deprivation and time on task (Wilkinson, 1970). As with the CAQ, given the length of the training sessions, the USRT was included in the test battery to permit an assessment of the extent to which fatigue was a factor in the study - if such an assessment was deemed to be necessary. In this computer-administered task, subjects are instructed to press the space bar as quickly as possible following the appearance of a red square on the computer screen. Stimuli are randomly presented at 1 to 10 second intervals. Sessions were 10 minutes in duration and a subject's score was the mean of their reaction time.

REFERENCES

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77*, 598-614.

Arthur, W., Jr. (1991). *Individual differences in the prediction and training of complex perceptual-motor skill tasks: The development and validation of the Computer-Administered Visual Attention Test.* Tech. Report No. 8. College Station, TX: Texas A&M University, Psychology Department.

Arthur, W., Jr., & Doverspike, D. (1992). Locus of control and auditory selective attention as predictors of driving accident involvement: A comparative longitudinal investigation. *Journal of Safety Research, 23*, 73-80.

Arthur, W., Jr., & Graziano, W. (1994). *Relationships among the Big Five personality dimensions and driving accident involvement*. Unpublished manuscript.

Arthur, W., Jr., & Olson, E. (1991). Computer attitudes, computer experience, and their correlates: An investigation of path linkages. *Teaching of Psychology, 18*, 51-54.

Arthur, W., Jr., Strong, M. H., Jordan, J. A., Williamson, J. E., Shebilske, W. L., & Regian. (in press). Visual attention: Individual differences in training and predicting complex task performance. *Acta Psychologica*.

Arthur, W., Jr., Strong, M. H., & Williamson, J. E. (1994). Validation of a visual attention test as a predictor of driving accident involvement. *Journal of Occupational and Organizational Psychology, 67*, 173-182.

Arthur, W., Jr., Young, B., Jordan, J. A., & Shebilske, W. (1994). *Effectiveness of dyadic training protocols: The influence of trainee interaction anxiety*. Manuscript submitted for publication.

Austin, J. T., Humphreys, L. G., & Hulin, C. L. (1989). Another review of dynamic criteria: A critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology, 42*, 597-612.

Avolio, B. J., Alexander, R. A., Barrett, G. V., & Sterns, H. L. (1981). Designing a measure of visual selective attention to assess individual differences in information processing. *Applied Psychological Measurement, 5*, 29-42.

Barrett, G. V., Alexander, R. A., & Doverspike, D. (1992). The implications for personnel selection of apparent declines in predictive validities over time: A critique of Hulin, Henry, and Noon. *Personnel Psychology, 45*, 601-617.

Barrett, G. V., Morris, S. B., & Alexander, R. A. (1993). The effect of factor analytic method on the interpretation of the changing abilities requirements model: An example of the tools-to-theory hypothesis. *Human Performance, 6*, 23-47.

Carrier, C. A. & Sales, G. L. (1987). Pair versus individual work on the acquisition of concepts in a computer-based informational lesson. *Journal of Computer-Based Instruction, 14*, 11-17.

Christal, R. E. (1976). *What is the value of aptitude tests?* Paper presented at the 18th Annual Conference of the Military Testing Association, Gulf Shores, AL.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1991). *The NEO PI/FFI Manual Supplement*. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Dambrot, F. H., Watkins-Malek, M. A., Silling, M. S., Marshall, R. S., & Garver, J. A. (1985). Correlates of sex differences in attitudes toward involvement with computers. *Journal of Vocational Behavior, 27*, 71-86.

Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology, 43*, 717-744.

Donchin, E. (1989). The learning strategies project: Introductory remarks. *Acta Psychologica, 71*, 1-15.

Folkard, S., Monk, T. H., & Lobban, M. C. (1978). Permanent night nurses. *Ergonomics, 21*, 785-799

Ford, J. K., Quinones, M., Sego, D., & Speer Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-527.

Foss, M. A., Fabiani, M., Mane, A. M., & Donchin, E. (1989). Unsupervised practice: The performance of the control group. *Acta Psychologica, 71*, 23-51.

Frederiksen, J. R. & White, B. Y. (1989). An approach to training based on principled task decomposition. *Acta Psychologica, 71*, 89-146.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.

Gopher, D. (1992). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Korenblum (Eds.), *Attention and performance XIV*: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience - A silver jubilee. Cambridge, MA: MIT Press.

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skill*, 33, 1335-1342.

Gopher, D., Weil, M., & Bareket, T. (1992). The transfer of skill from a computer game trainer to actual flight. *Proceedings of the Human Factors Society 36th Annual Meeting*, 1285-1290.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147-178.

Gottesfeld, N., & Arthur, W., Jr. (1994). *Openness and conscientiousness as predictors of performance on a complex perceptual-motor skill task*. Unpublished manuscript.

Hanges, P. J., Schneider, B., & Niles, K. (1990). Stability of performance: An interactionist perspective. *Journal of Applied Psychology, 75*, 658-667.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin, 107*, 328-340.

Jones, W. H., Briggs, S. R., & Smith, T. G. (1986). Shyness: Conceptualization and measurement. *Journal of Personality and Social Psychology, 46*, 629-631.

Kyllonen, P. C., Christal., R. E., Woltz, D.J., Shute, V. J., Tirre, W. C., & Chaiken, S. (1990). *Cognitive abilities measurement (CAM) test battery: Version 4.0* [Computer program]. Brooks Air Force Base, TX: Armstrong Laboratory.

Kelly, C. R. (1969). *Manual and automatic control.* New York: Wiley.

Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of Personality Assessment, 47*, 66-75.

Leary, M. R. (1986). The impact of interactional impediments on social anxiety and self-presentation. *Journal of Experimental Social Psychology, 22*, 122-135.

Leary, M. R., Atherton, S. C., Hill, S., & Hur, C. (1986). Attributional mediators of social inhibition and avoidance. *Journal of Personality, 54,* 188-200.

Leary, M. R., & Kowalski, R. M. (1987). Manual for the Interaction Anxiousness Scale. *Social and Behavioral Sciences Documents, 16*(MS. No. 2774), 2.

Lintern, G., & Gopher, D. (1978). Adaptive training of perceptual motor skills: Issues, results and future directions. *International Journal of Man-Machine Studies, 10,* 521-551.

Logg, E. (1993). *Microsoft Arcade: Asteroids [Computer program].* Redmond, WA: Microsoft Corporation.

Mane, A. M., Adams, J. A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, 71,* 179-196.

Mane, A. M. & Donchin, E. (1989). The Space Fortress game. *Acta Psychologica, 71,* 17-22.

Mihal, W., & Barrett, G. V. (1976). Individual differences in perceptual information processing and their relation to automobile accident involvement. *Journal of Applied Psychology, 61,* 229-233.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space Fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta-Psychologica, 71,* 243-257.

Raven, J. C., Court, J. H., & Raven, J. (1985). *A manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis.

Shebilske, W. L., Jordan, J. A. & Arthur, W., Jr. (1993). Cognitive factors in automated instruction for individuals and groups [Abstract]. *Proceedings of the 34th Annual Meeting of the Psychonomic Society, 1,* 24.

Shebilske, W. L., Jordan, J. A. Arthur, W., Jr., & Regian, J. W., (1993). Combining a multiple emphasis on components protocol with small group protocols for training complex skills. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society, 2,* 1216-1220.

Shebilske, W. L., Regian, J. W., Arthur, W., Jr., & Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34,* 369-374.

Shebilske, W. L., & Regian, J. W., (1992). Video games, training, and investigating complex skills. *Proceedings of the 36th Annual Meeting of the Human Factors and Ergonomics Society, 2,* 1296-1300.

Synder, M. (1987). *Public appearances, private realities: The psychology of self-monitoring.* New York: W. H. Freeman.

Strong, M. H. (1992). *An assessment of the criterion-related validity and improvements in the utility of the Computer Administered Visual Selective Attention Test.* Unpublished masters thesis, Texas A&M University, College Station.

Theurer, D. (1993). *Microsoft Arcade: Tempest [Computer program].* Redmond, WA: Microsoft Corporation.

Wilkinson, R. T. (1970). Methods for research on sleep deprivation and sleep function. In E. Hartman (Ed.) *Sleep 1980: Fifth European Congress of Sleep Research.* Basel: Karger 1981.

Wilkinson, R. T., & Houghton, D. (1982). Field test of arousal: A portable reaction timer with data storage. *Human Factors, 41,* 487-493.